Search in Social Networks

Roby Muhamad

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Science

COLUMBIA UNIVERSITY
2010

ABSTRACT

Search in Social Networks

Roby Muhamad

More than four decades ago, Stanley Milgram and his collaborators performed a series of original experiments to test the small-world hypothesis: whether any random pair of individuals can be connected through short chains of acquaintances. They found support for the hypothesis and their results are currently known as the “six degrees of separation.” Closer examinations, however, revealed that Milgram’s experiments actually confirmed two related but distinct hypotheses: topological and algorithmic small-world hypotheses. Topological small-world hypothesis posits that there are short paths connecting two individuals. Algorithmic small-world hypothesis asserts that individuals with limited information can actually find these short paths by actively searching social networks. The goals of this dissertation are two-fold: (1) to test the algorithmic small-world hypothesis, and (2) to understand the mechanisms that make the algorithmic small-world possible. To achieve the first goal, we used data from our global internet-based search experiment and, using a novel statistical method, estimated algorithmic distance distributions. Then we used computational models to understand search processes and identified search strategies, individual characteristics, and structural conditions that increase the probability of success in search processes.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This dissertation is my greatest intellectual adventure thus far, of which a number of people have helped me along the way. I thank my adviser Duncan Watts who guided and supported me from the very beginning, his intellectual influence on my thinking cannot be overstated. I am also indebted to Peter Bearman for his insights and encouragement, and for showing me the beauty of sociology. I am grateful to Josh Whitford, Bruce Kogut, David Stark, Harrison White, and Andrew Gelman who provided useful comments for this dissertation.

Peter Dodds was a constant source of encouragement, advice, and inspiration during his stay at Columbia. Sharad Goel was responsible for bringing such mathematical rigor in Chapter 3, I thank him very much. Peter Hausel, Alison Binkowski, and Rob Stewart helped me in creating and maintaining the small-world project website. I am also grateful to John Lee whose algorithm for generating networks I used in Chapter 4.

I thank the many support and advice from students at the Sociology Department and former members of the Collective Dynamics Group, also the staff at the Sociology Department and ISERP, who were very supportive.

Lastly, I thank my friends in NYC who were always around to share the emotional burden of writing this dissertation, reminding me of how lucky I am to be able to write this dissertation. Finally, to Tika and Toby whose presence makes everything worthwhile, to mamah and papah who were there from the very beginning.

To Tika

CHAPTER ONE: INTRODUCTION

Men are conjoined by a vast network of acquaintanceship. Brown knows Jones, Jones knows Robinson, etc.; and by choosing your farther intermediaries rightly you may carry a message from Jones to the Empress of China, or the Chief of the African Pigmies, or to anyone else in the world…

—William James, *Pragmatism*

1.1 Topological and algorithmic small-world hypotheses

More than four decades ago, Stanley Milgram and his collaborators performed a series of experiments and showed that, contrary to our common sense about (pre-Internet) modern society and a multitude of social barriers, anyone can reach anyone else through the average of six intermediaries. This finding—later known as the "six degrees of separation" idea—is more formally known as the small-world hypothesis. This dissertation is a study of the small-world hypothesis and tries to answer two questions: (1) What is the scope of the small-world hypothesis and the evidence supporting it? (2) What are the implications of small-world structure for individuals? Particularly, we focus on how individuals can exploit indirect short paths to obtain resources or information embedded in their social networks.

To answer both questions in a meaningful way, we need to differentiate topological and algorithmic aspects of the small-world hypothesis. The *topological* small-world hypothesis claims that for any pairs of individuals we can construct a short path connecting them, where "short" usually means that the length of the path is proportional to the logarithm of the population (Watts and Strogatz 1998). The *algorithmic* small-world hypothesis asserts a stronger claim. In addition to the presence of short paths, the algorithmic small-world hypothesis

claims that ordinary individuals with limited information can find those short paths (Kleinberg 2000). Consequently, we can construe two kinds of connections and distances in social networks. Topological connections among individuals are network connections that are not necessarily known to individuals. In other words, individuals are not always aware of the topological distance connecting them to other individuals. In contrast to the topological connections, algorithmic connections are social networks connections that individuals have awareness of. The algorithmic distance is the resulting distance when an individual actively searches social networks to find another individual.

The distinction between topological and algorithmic small-world hypotheses is also useful because each hypothesis corresponds to a different social process (Watts 2003). The topological small-world hypothesis focuses on whether or not short paths exist regardless of whether individuals are aware of them. Contagion or diffusion-type processes come to mind as examples of how the topological small-world hypothesis could determine the dynamic of the spread. In a contagion, individuals do not actively seek to be infected, and their risk of getting infected is mostly related to how they are topologically connected to infectives. On the contrary, in a search process such as "networking," a person actively seeks a connection to a particular person. Thus, algorithmic connectivity is necessary because individuals not only are connected through short paths to target persons, but also they need to be able to find these short paths.

The differentiation between the topological and algorithmic small-world hypotheses implies that we need different kinds of empirical evidence to validate

them. To show that a network is connected in the topological sense, what is needed is to measure the average path length for a given network. A large number of studies have produced consistent findings across different kinds of networks that the average path length is proportional to the logarithm of the network size. For example, online communication networks (Kossinets and Watts 2006; Leskovec and Horvitz 2008), organizational networks (Adamic and Adar 2005; Kogut and Walker 2001), and biological and technological networks (Watts and Strogatz 1998) have been shown to follow the topological small-world principle.

On the other hand, evidence for the algorithmic small-world hypothesis is more limited. The evidence must include search processes in which individuals successfully navigate their social networks to obtain resources, information, or services. In one study of a naturally-occurring search process, Granovetter (1995) studied the process of getting a job. One well-known result from this study is that weak ties play an important role in getting information about a job (Granovetter 1973). However, Granovetter's results also show that about 84% of respondents received job information directly from the prospective employer or through at most one intermediary; in fact, the maximum length of chains in this study was four (Granovetter 1995). In another study of a natural search process, Lee (1969) conducted post-hoc interviews with women who had had an abortion about how they found an abortionist. At that time, abortion was still illegal in the United States and hence women who were looking for an abortionist had to rely on informal channels such as their social networks. Lee found that about 12% of

the women found an abortionist through a chain of nine or more intermediaries, and the median of chain length was five. According to the study, about 61% of women had to restart their searches several times, exploring different contact channels.[1]

These case studies, while showing successful instances where individuals have the ability to navigate networks, by no means provide the evidence that social networks are generally searchable. From these results, it is still not clear whether ordinary individuals can locate any other individuals as suggested by the algorithmic small-world hypothesis because these two case studies could be just special cases. In other words, while case studies of naturally-occurring search processes contribute to our understanding of substantive social scientific problems, the lack of control in case studies makes it difficult to tease out the general mechanisms of the search process and the underlying network architecture that makes search possible. Therefore, so far, the only direct empirical evidence for the small-world hypothesis comes from experiments that use Milgram's small-world method, which we will review in the next section.

1.2 Evidence for algorithmic small-world hypothesis

Research on the small-world problem originated in the early 1950s when political scientist Ithiel de Sola Pool and mathematician Manfred Kochen started a pioneering research project that dealt with the quantification of social

[1]Based on her interviews, Lee (1969) found that women who had had legal hospital abortions (either because they had valid medical conditions, missrepresented or exaggerated their actual conditions, or traveled to a country where abortion was legal) reported higher satisfaction and fewer complications than women who went to illegal abortionists. This raises the question of the quality of results from search in social networks, which is still understudied.

structures. The journal *Social Networks* published their findings much later on as part of the first issue (Pool and Kochen 1978). Pool believed that “the very stuff of politics” is about exerting influence through social contacts, and together with Kochen, they formulated the problem of political access as the process of finding a chain of contacts that led to someone with political power. They examined the probability that two individuals, chosen at random, would have common friends, commonly known as the “small world of the cocktail party”[2]. Pool and Kochen also calculated the probability of how two random people, who did not share mutual friends, could be connected through chains of acquaintances. Their model, however, assumed that each person had the same number of contacts whereby each person’s contacts are entirely distinct from everyone else’s. The fact that there are usually many shared contacts amongst individuals rendered the Pool and Kochen model impractical.

Stanley Milgram (1967) picked up where Pool and Kochen left off. He empirically investigated the average number of intermediaries of acquaintances that are needed to connect two randomly-chosen people. Milgram invented a simple method that would later be known as the small-world method (Milgram 1967; Travers and Milgram 1969).[3] Milgram assumed that the actual process of establishing a link between two individuals traveled in one direction, from initial senders to a target person. In conducting his experiment, Milgram recruited a

---

[2] In their paper, Pool and Kochen also had a lengthy discussion about the estimation of acquaintance volume because they thought that the number of contacts is proportional to the level of (political) influence one has.

[3] Before Milgram, Rapoport and Horvath (1961) conducted an empirical study on connectivity of social networks, albeit the method is different from Milgram and was conducted in a small school population.

random sample of people to act as starters. All of them received basic information about a target person, and each was asked to forward a message toward the target person. If they knew the target, then they sent the message directly to the target, but if they did not know the target, then they chose acquaintances who they thought could bring the message closer to the target. On receiving the message, each acquaintance was to repeat the process until the message reached the target.

Using this method, Milgram conducted two experiments. In the first experiment, Milgram choose the wife of a Divinity School student in Cambridge, Massachusetts as the target, while he selected random individuals from Wichita, Kansas as starters. For the second experiment, Milgram selected a stockbroker in Sharon, Massachusetts as the target, while he based his selection of individuals as starters on these three categories: blue chip stockholders from Omaha, random participants from Omaha, and random participants from Boston, Massachusetts.[4] The two experiments are known as the “Kansas study” and the “Nebraska study” respectively (Milgram 1967).

In the Kansas study, only 3 (6%) of the 50 chains started reached the target, where completed chains required an average of eight intermediaries. Of the 296 starters selected for the Nebraska study, 100 were blue chip stockholders, 96 were randomly-selected individuals from Omaha, while 100 were arbitrarily-chosen individuals from Boston. Within the blue chip stockholders

---

[4]Kleinfeld (2002) pointed out that these groups of participants were hardly random for a target who is a stockbroker living in Sharon, Massachusetts. In addition, sources in the Kansas study were recruited through mailing lists so that biased sampling toward high-status people and newspaper advertisements were framed in a patriotic theme: “Could you, as a typical American, contact another citizen, regardless of his walk of life?”

group, 78 out of the 100 messages actually got started and 24 of them reached the target in Sharon, Massachusetts. For the random group originating in Nebraska, out of the 76 messages sent by initial senders, 18 of them completed the chain. As for the participants from Boston, 63 messages got started and 22 of them reached the target. In summary for the Nebraska study, there were 296 starters; from these starters, 217 passed the message, and 64 (30%) of them reached the target. The average number of intermediaries between starters and targets was 5.2.

Both of Milgram's experiments yielded relatively low completion rates. However, follow-up studies that used smaller sample sizes than Milgram's experiments showed higher completion rates. For example, in a business firm (Lundberg 1975), a university (Shotland 1976), and a city (Guiot 1976) study, the completion rates were 57%, 69%, and 85% respectively. Thus, whereas small-world experiments in large populations yielded short-length successful chains with low completion rates, experiments in small populations had higher completion rates.

A low rate of chain completion also appeared in Korte and Milgram's study of acquaintance networks between racial groups (Korte and Milgram 1970). They chose 18 targets in New York City, half of them white and half black, and 540 white starters from Los Angeles. Half of the starters were given white targets and the other half received black targets. None of these starters had any information about the race of the target persons; hence, bias from racial prejudice could be disregarded, but there was a bias toward assuming the target was white. The

completion rate was 33% for white targets and 13% for black targets. Chains only crossed racial groups at the last link, typically from a white superior to a black subordinate who was the target. Their findings also indicated that there was no significant difference in mean path length: 5.5 intermediaries for completed white chains and 5.9 for completed black chains. Another study of communication between racial groups in a localized urbanized area involved 298 volunteers, and 30% of the packages reached the target person (Lin, Dayton and Greenwald 1978). Lin et al. found that participants were less willing to send messages across racial boundaries and that the search process was more effective in the direction from high-status to low-status people. However, the study also found that lower-status people had a strong desire or willingness to reach higher-status targets. These studies seemed to indicate that social status could negatively impact the completion rates.

There are at least two shortcomings from the previous small-world experiment. The first is that subsequent small-world experiments after Milgram were conducted in smaller populations than the original experiment (we will review previous small-world experiments in more detail in the next section); thus, it is not clear whether the finding still holds for larger populations. In addition, recent developments have made the world even more connected, but the gap between social statuses is wider than when Milgram and his collaborators did their experiments. Thus, there is a need to replicate small-world experiments in the current setting on a scale that is larger than the original experiment so that we are able to determine the scope of the algorithmic small-world hypothesis.

Another shortcoming is that even though there are chains that reach their intended targets within short steps, the bulk of the chains never reach their targets. We call this problem the attrition problem. The presence of attrition makes it difficult to interpret findings from small-world experiments; hence, the evidence for the algorithmic small-world hypothesis is based only on small portions of data points. Part of this dissertation is dedicated to address these two shortcomings by conducting a global small-world experiment and a deeper investigation of the attrition problem.

1.3 Network structure and individual networking

Social capital theory asserts that social relationships are a form of capital from which individuals can obtain resources. These resources can be either in the form of social norms of mutual obligations and expectations that are generated by cohesive social ties, or in the form of information possessed by individuals embedded within social networks (Bourdieu 1986; Coleman 1988; Lin 2001). Here, we focus on the process of information searching to access social capital. The process of information searching is relevant in a wide range of social processes: the traditional bazaar (Geertz 1978), getting a job (Bearman 2005; Granovetter 1995), searching for an abortionist (Lee 1969), entrepreneurs searching for exchange partners (Hoang and Antoncic 2003; Sorenson and Stuart 2008), and organizational problem-solving (Sabel 2004; Singh, Hansen and Podolny 2008).

One may wonder why conducting searches in social networks is still necessary when technology such as Internet search engines, social networking sites and knowledge management systems are widely available. There are at least two reasons why information search through social networks is still desirable. The first reason is that intermediaries can help identify targets and give signals about targets' quality and reliability. Therefore, networking serves as an uncertainty-reduction mechanism that could result in a better match between source and target compared to asocial searches (for example, see Andrew, Markus and Barry 2006). The second reason is that knowledge databases are useful only when the information that we are looking for is known and has been codified and entered into the system. If the information or expertise that we are looking for is less tangible or novel, then target identification can only be made by individuals who have faced the same or similar problems previously. This is especially true in knowledge-intensive industries (Singh, Hansen and Podolny 2008) or illicit searches (Lee 1969). Here, intermediaries are important not only for connecting to targets but also for identifying targets themselves.

At its most general level, the search process in small-world experiments is the same as networking activities in the real world in the sense that both are based on the premise that individuals can traverse their social networks to find target persons. The similarity, however, ends there. The search process in small-world experiments is artificial because the search is designed for a particular purpose in the context of a controlled experiment; nobody in the real world conducts searches the way participants in the experiment do.

To elaborate further, it is useful to create a typology (Table 1.1) of searches and situate the kinds of searches that are studied here within this typology. We can make a classification using two dimensions. The first dimension is whether targets are known or unknown, and the second dimension is whether we have a single individual or a group of individuals as targets. Small-world experiments are examples of the search process with a specific target person whose identity is known (quadrant I, Table 1.1). In reality, however, we rarely do this first type of search. Instead, searches in real life are mostly involved with targets as a group or type of people or things whose identities are known (quadrant III, Table 1.1). Looking for an apartment or a job is an example of type III searches; other examples would be searching for a specialist doctor or a specific service provider. Real-world searches also include targets that are not known to searchers, i.e., searchers do not know what they are looking for and will recognize the target only when they find it. When the target is an unknown single individual (quadrant I, Table 1.1), we have a type of search that is similar to a reporter pursuing a secret would-be informant. Multiple unknown targets (quadrant IV, Table 1.1) are common when we individuals need to solve novel problems such as when doing research or when organizations are adapting and exploiting an uncertain environment (Stark 2009). In this dissertation we will focus on searches with known individual and collective targets (quadrant I & III Table 1.1).

	Known targets	Unknown targets
Single target	I e.g., small-world experiment	II e.g., secret informant, a suicide bomber
Multiple targets	III e.g., apartments, jobs, doctors, investors	IV e.g., research, innovation

Table 1.1 Typology of search.

To explore how individuals perform networking to find a type of known target, we construct a computational model that is described in chapter 4. We extend a network formation model that was first proposed by Watts, Dodds, and Newman (2002) and incorporate a model of networking activities. The goal is to understand factors that hinder individuals from networking successfully and explore ways to overcome this problem.

A note about our approach in this dissertation. The approach follows the strategy of analytical sociology (Hedstrom 2005; Hedstrom and Bearman 2009). According to this point of view, explanations for a social fact must refer to its micro foundations: how individual actions and their relations together produce collective outcomes. In the context of our problem, this general problem is translated into the problem of understanding the origin of the discrepancy between the topological and the algorithmic distances. To achieve this goal we used a multitude of approaches: we deployed a controlled experiment to test a

hypothesis, built statistical models to understand statistical associations among individual attributes and relationships to understand the outcome of the experiment, and constructed a generative model to tease out relevant mechanisms that bring about observed outcomes. We hope our results, then, will serve as a guide for further studies that can isolate the details of why the algorithmic distance could become very different from the topological distance. Therefore, the analytical strategy would render the accumulation of knowledge possible.

1.4 Dissertation outline

In Chapter 2, we will describe the experimental design and analyze some results from the experiment. We then discuss one particular problem, the attrition problem, in Chapter 3. The attrition problem comprises two aspects: (1) how to model attrition as functions of variables that are recorded in the experiment, and (2) how to construct an unbiased estimate of chain length and simultaneously incorporate heterogeneity in attrition that is captured by the attrition model. We continue in Chapter 4 by constructing a computational model to delineate various search mechanisms and their effect on search outcomes. Finally we offer our conclusions in Chapter 5.

CHAPTER TWO: A GLOBAL SEARCH EXPERIMENT[5]

More than forty years have passed since Milgram’s experiment and the idea of “six degrees of separation” has become part of our popular culture. Nevertheless, the empirical basis of the assertion that anybody is only six steps away from anybody else came only from Milgram’s original large-scale experiment. Furthermore, only 146 randomly-picked individuals were "distantly associated" from the target; the rest of initial senders were either closely related to the target’s occupation or lived in the same city as the target. Additionally, only 21 were successful chains. Thus, the six degrees of separation claim is actually supported by merely 21 data points. There is a need to test the small-world hypothesis using a large-scale experiment that is at least comparable or even larger than Milgram’s original experiment. It is also interesting to test whether Milgram’s findings still hold in a world of increasing connectivity in a context of increased inequality. We try to resolve these problems by conducting a global search experiment that uses Milgram’s small-world method, which will be the focus of this chapter.

2.1 Experimental procedure

We collected our data through a global social search experiment using Milgram’s small-world method (Dodds, Muhamad and Watts 2003). Instead of using the postal service as Milgram did, we used e-mails, which allowed access

[5] Some of the results in this chapter have been reported in Dodds, Peter S., Roby Muhamad, and Duncan J. Watts. 2003. "An Experimental Study of Search in Global Social Networks." *Science* 301:827-829.

to a more diverse and larger population with minimal cost (Best, Krueger and Smith 2001). Furthermore, the number of Internet users in the world has grown from around 0.4% of the world population in 1995 to almost 24% in 2008 (Stats 2008).

Participants came to our website, registered and entered their basic demographic information. For each participant, we randomly selected a target and displayed the target's information. We asked senders to choose the next person in the chain to be someone whom they knew, and who was "closer" to the target. Senders could only select one contact person for each target. In the next step, senders entered the name and e-mail address of the message's recipient. We also asked them to provide the type, strength, and origin of their relationship with the recipient. Senders could add a short personal message for the recipient if they wanted to. All messages were sent through our centralized server (http://www.smallworld.columbia.edu), thereby allowing us to track the progress of all messages and discourage "unofficial" messages circulating around, even though we could not prevent people from forwarding messages without using our website.

Anyone could participate in this project, whether volunteering as a starter or a target. Our website consisted of two sections: the public section, which contained information about the project and the small-world problem in general, and the password-protected section, in which a participant entered their personal information, selected a contact, sent a message and tracked the progress of the message. In the public section, we also provided sample pages of what

participants would typically see after they logged in to familiarize them with the sending process. To participate in this project, a volunteer visited our website and registered by supplying us with her name and e-mail address so that we could reply with an e-mail containing the login ID and password required to enter the sending site. The E-mail also included information about the target person, a short explanation of the project, and a step-by-step guide to participate.

After participants sent messages, they would be asked whether they wanted to participate again in another chain with a different target person. If they chose to do so, a new sending page with a new target person would be displayed. If not, they were directed to the public section. Another way for a participant to obtain a new target person was by registering again, prompting the computer to randomly assign them a different target. When participants chose to get a new target person, they did not have to enter their personal data again.

The new participants (who were referred to by previous senders) then received an e-mail from us with the name of the previous participant shown as the sender of the e-mail. This e-mail contained the target person's information, a login ID, a password and a personalized message from the sender, intended to reduce the tendencies of recipients to delete this message as junk e-mail. Then, the new participants were instructed to follow the same steps as described above with one exception; they must verify that they knew the sender by recognizing the sender’s name and e-mail address. This verification was important, since the only requirement in selecting the contact person was that the sender knew the

person personally. To assure this, the contact person was required to verify that they knew the sender.

Participants who did not respond in one week received a courtesy e-mail from the system automatically to remind them that they had messages to send. In this reminder e-mail, we gave them three options, each with its own direct URL so that they could click the links to indicate their choice. It also notified them that they had one week to respond. So in total, they had two weeks to respond. These options were:

1. Participate. This option was for those who wanted to participate but forgot or had not participated.
2. Participate but need more help. This link was for those who wanted to participate but were not sure what to do. The link would bring them to a page describing the procedure again. We also provided a dialogue box so participants could write the reason why they thought the experiment was too hard. We also provided an option if they wanted to get a new target person if they thought the current target person was too hard to find.
3. Remove. This option was for those who did not wish to participate in the small-world project and did not want to receive any further messages about the project; it was an unsubscribe option. We also asked why they did not want to participate in the experiment.

If after receiving the reminder the participants still did not reply in one week, the system would terminate the participant and send e-mails to the prior

message-holders informing them that their contacts had failed to respond. They would then be offered another chance to resend the message to a different person. We implemented this mechanism as an effort to keep the chains alive. We limited this backward activation of senders to only one step.

When participants received a message, theoretically, they might choose to send the message to all of their friends to get the message closer to the target, but we restricted each participant to only be able to send one message for each message they received. We restricted the message sent by participants for two reasons: first, not only were we interested in the average path length, we were also interested in how people chose these paths. To address both questions, it was necessary to limit the number of messages a participant could forward, forcing them to use their optimal strategy. The second reason was to avoid exponential growth of the number of messages circulating on the Internet. Computer worms were very dangerous for local and global networks; exponential growth could easily cripple our server or even cause problems to the Internet.

All participants, including targets and senders, were unpaid volunteers and we did not offer any incentives for participating or completing the chains. Thus, to get volunteers, we relied heavily on media reports of the project that listed a direct link to our website. In addition to conventional media, such as newspapers and radio broadcasts, information about our project has also spread in the cyber-world through websites, mailing lists, cyber-forums and chat rooms.

We obtained a response rate of 35%, which was higher than was typical for e-mail surveys (Salo et al. 2000). Unfortunately, as more people use e-mail as

a relatively easy way to access a greater number of people with diverse characteristics, e-mail communications have been plagued by the widespread circulation of unsolicited e-mails (junk e-mails or spam), computer viruses and worms. Although the response rate for Internet surveys was higher than that for postal surveys (the average response rate for mail surveys was between 1% and 2% (Nucifora 2002)), the general trend was that the response rate for Internet surveys was decreasing dramatically (Sheehan 2001). We have anecdotal evidence that automated spam filters blocked our messages, so that willing individuals mistook our messages as commercial spam. The problem of unsolicited e-mails and mail was so pervasive that it was unlikely that we could have achieved the 75% response rate that Milgram and Travis accomplished in the '60s.

The task for participants was to send e-mail messages toward a randomly-assigned target person chosen from eighteen available targets (Table 2.1). Initially we used commercially-available e-mail lists to recruit potential senders. Although this method was unsuccessful, with a response rate of less than 0.5%, we did gain the attention of the global media (e.g., *The New York Times*, *Newsweek*, *US News and World Report*, *CNN*, and *BBC*). This allowed us to switch to a passive recruitment process where initial senders came and signed up on our website after learning about the experiment from various offline and online media. In order to get as many participants as possible, we did not control the characteristics of senders. The first six target persons were acquaintances of the research team (half of them in the United States and half outside the United

States). The other twelve targets came later through solicitation from the experiment website (they were chosen from about four thousand volunteers, with the aim of creating a rich and diverse selection of targets). All targets provided their full names, cities, states or provinces, countries of residence, current occupations, and educational history. Some targets also provided their age and occupational histories.

Target	City	Country	Occupation	Gender	$r(r_0)$	*N*	$N_c(\%)$	<L>
1	Novosibirsk	Russia	Ph.D. student	F	.64(.42)	8907	36(.4)	4.5
2	New York	USA	Journalist	F	.64(.37)	6495	33(.5)	3.5
3	Bandung	Indonesia	Grad. Student	M	.66(.44)	8759	0	n/a
4	New York	USA	Editor	F	.62(.35)	6150	57(0.9)	4
5	Ithaca	USA	Professor	M	.59(.35)	6411	212(3.3)	4
6	Melbourne	Australia	Travel Consultant	F	.61(.38)	6106	24(0.4)	5.2
7	Sortland	Norway	Veterinarian	M	.65(.42)	4650	18(0.4)	4.2
8	Perth	Australia	Policeman	M	.65(.41)	4831	8(0.2)	5.1
9	Omaha	USA	Insurance Agent	F	.67(.46)	4848	4(0.1)	5.8
10	Welwyn Garden	UK	Retired	M	.68(.42)	7041	1(0.01)	4
11	Paris	France	Librarian	F	.65(.39)	4509	3(0.1)	5
12	Talinn	Estonia	Archival Inspector	M	.66(.46)	4835	10(0.2)	4.2
13	Munich	Germany	Journalist	M	.63(.39)	4696	42(0.9)	4.8
14	Split	Croatia	Student	M	.66(.45)	7051	0	n/a
15	Gurgaon	India	Technology Consultant	M	.68(.43)	4846	15(0.3)	3.6
16	Managua	Nicaragua	Analyst	M	.69(.47)	6942	3(0.04)	5
17	Katikati	New Zealand	Potter	M	.64(.39)	4439	13(0.3)	4.3
18	Elderton	USA	Pastor	M	.67(.42)	4779	12(0.3)	4.3
Totals					*.65(.41)*	*106,295*	*491(0.5)*	*4.2*

Table 2.1. List of targets. Average and initial attrition rates denoted by r and r_0 respectively. *N* is the number of chains started to reach the corresponding target. N_c is the number of chains that reached the targets, and *<L>* is the mean path length of completed chains.

2.2 Analysis

Before we continue to the results of our experiment, we note that although the findings reported here are based on the same experiment that was used in the earlier publication (Dodds, Muhamad and Watts 2003), changes in our coding and some subsequent cleaning of the data resulted in somewhat different figures

for number of participants and completed chains. There are a number of reasons for these discrepancies. First, we included a number of chains that were originally excluded in Dodds et al. (2003) because they contained individuals who could not be identified previously because of database error. Second, Dodds et al. (2003) required actual e-mails to be sent between two people in order to establish a connection. After closer examination, however, we found that there were people—especially those directly adjacent to targets—who received more than one message but who had continued only one message. Thus, we counted a connection whenever we knew that it existed from previous e-mail exchanges; thus we had higher total numbers for both incomplete and completed chains. Finally, we have added three demographic variables (ethnicity, work industry, and work position) that were inaccessible previously due to many participants answering the “other” category. We solved the problem by checking manually all uncategorized answers and then putting them into relevant categories.

In total, 98,865 individuals from 168 countries registered at our website, and initiated 106,295 chains toward the eighteen targets in thirteen countries (Table 2.1). An individual could only send one message for each target, but could participate in multiple chains with different targets, resulting in a greater number of chains than the number of senders. More than half of our participants came from North America, and their characteristics resembled the typical characteristics of individuals with access to the Internet (Chen, Boase and Wellman 2002). They were predominantly young, college-educated, white, Christian, and middle-class professionals (Figure 2.1). A total of 491 (0.5%)

chains successfully reached their targets with the average length of 4.2 steps. The average attrition rate across all targets was 65%, but the initial attrition rate was 41%. The initial attrition rate was lower because it was the rate of attrition for participants who volunteered to initiate chains and so they were self-selected. Target number five, who is a professor in a large university in the Northeast, had the lowest attrition (59%) and thereby the highest completion rate (3.3%). Two targets, students in Indonesia and Croatia, never received a message.[6]

[6]The student in Indonesia, however, reported that he received a message by phone and so it was not recorded; he tried unsuccessfully to help the contact to send the message by e-mail. The target in Paris also reported that, on several occasions, she had to personally help her contact to send her a message by e-mail. These cases illustrate how the medium used could affect chain progression. For the case of the Croatian student, we were unable to contact him again after he signed up to become a target.

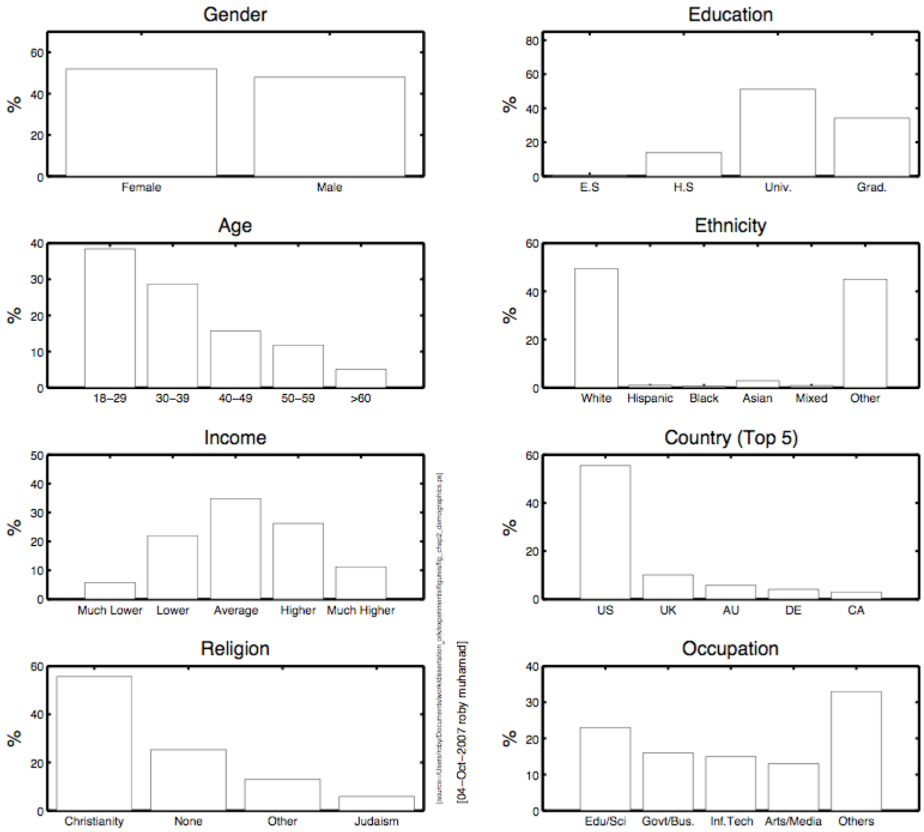

Figure 2.1 The demographic characteristics of participants. To maximize participation, some questions were voluntary. Response rates for these questions were as follows: Income (64%); Education (79%); Occupation (86%); Age (87%); Religion (69%); Ethnicity (81%).

When a participant sent a message to a contact, we asked how he or she had come to know the contact person and the type and strength of their relationship (Figure 2.2). Participants mostly used friendship ties in preference to business or family ties when sending messages. Yet, these friendship ties were mostly formed through business and school affiliations. People favored someone whom they considered “fairly close” when choosing the next person in a chain. Thus, the most useful relationship for sending messages was medium-strength friendships that originated in the workplace. Also note that most relationships used in this experiment originated from offline interactions; only about 7% of ties

originated from the Internet. Thus, our experiment was actually about networks of acquaintances, not just electronic social networks, and our use of e-mail was simply a tool to trace this network of acquaintances.

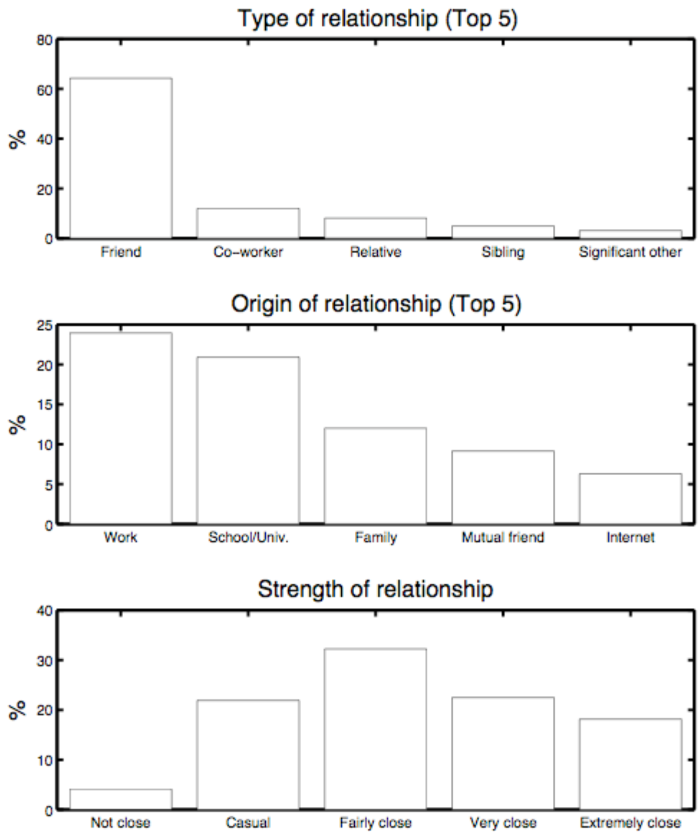

Figure 2.2 The type, origin, and strength of social ties used to direct messages. For types and origins of relationships, only the top five categories are listed.

To understand better the efficacy of social ties in directing messages, we compared completed and incomplete chains in terms of four relational variables: the origin, type and strength of relationship, and the reason for choosing the next recipient. We found that for all relational variables, categories used by participants for complete and incomplete chains are different ($p - value < 10^{-10}$,

standard χ^2 test). To discern which categories favored complete chains, we performed the following detailed analysis:

We start with the type of relationship variable that captures answers to the question "What is the nature of your relationship? This person is my…." The result is depicted in Table 2.2. Subscripts c and i correspond to complete and incomplete chains respectively. N is the frequency of each category; f is the relative frequency of each category; $\Delta = f_{c,x} - f_{i,x}$ is the absolute difference in relative frequencies between complete and incomplete chains where x is the index for categories; $\delta = 100(f_{c,x} - f_{i,x}) / f_{i,x}$ is the corresponding relative difference; *Rank* orders the categories by decreasing $|\delta|$, i.e., Rank 1 corresponds to highest value of δ. Except for N, all quantities are percentages. Categories are listed in order of increasing Δ; categories with higher Δ are more likely to be found in completed chains. As we have seen, participants used friendship ties extensively both in complete and incomplete chains. Yet, professional ties were disproportionately favored over familial and friendship ties in successful chains.

Type of relationship	N_i	N_c	f_i	f_c	Δ	δ	*Rank*
Friend	*53614*	*837*	*65.2*	*52.8*	*-12.4*	*-19.0*	*9*
Relative	*6837*	*58*	*8.3*	*3.7*	*-4.7*	*-56.0*	*11*
Sibling	*4185*	*35*	*5.1*	*2.2*	*-2.9*	*-56.6*	*12*
Spouse/Significant other	*2693*	*37*	*3.3*	*2.3*	*-0.9*	*-27.3*	*10*
Child	*585*	*3*	*0.7*	*0.2*	*-0.5*	*-71.4*	*13*
In-law	*464*	*7*	*0.6*	*0.4*	*-0.1*	*-16.7*	*8*
Parent	*859*	*15*	*1.0*	*0.9*	*-0.1*	*-10.0*	*6*
Junior	*122*	*2*	*0.1*	*0.1*	*-0.0*	*0*	*7*
Service provider	*425*	*15*	*0.5*	*0.9*	*+0.4*	*+80*	*4*
Senior	*334*	*17*	*0.4*	*1.1*	*+0.7*	*+175.0*	*1*
Other	*1816*	*52*	*2.2*	*3.3*	*+1.1*	*+50.0*	*5*
Client	*595*	*29*	*0.7*	*1.8*	*+1.1*	*+157.0*	*2*
Co-worker	*9732*	*478*	*11.8*	*30.2*	*+18.3*	*+155.1*	*3*

Table 2.2 Comparison of the type of relationship in complete and incomplete chains. Subscripts c and i correspond to complete and incomplete chains respectively. N is the frequency of each category; f is the relative frequency of each category; $\Delta = f_{c,x} - f_{i,x}$ is the absolute difference in relative frequencies between complete and incomplete chains where x is the index for categories; $\delta = 100(f_{c,x} - f_{i,x})/f_{i,x}$ is the corresponding relative difference; $rank$ orders the categories by decreasing $|\delta|$, i.e., Rank 1 corresponds to highest value of δ. Except for N, all quantities are recorded as percentages. Categories are listed in order of increasing Δ.

For the question "How did you get to know them?" results are shown in Table 2.3 where categories are ordered according to increasing Δ; all quantities are defined the same as in the previous analysis (Table 2.2; ties in successful

chains are much more likely to have formed in professional and educational settings. With respect to the question “How well do you know this person?”, weak ties constituted a disproportionate part of successful chains' weak ties, particularly casual ones (Table 2.4). Lastly, responses to the question “Why did you select this person to receive the message?” are depicted in Table 2.5. In successful chains, “similar profession” as the target was chosen 331% more frequently than in unsuccessful chains.

How initially met acquaintance	N_i	N_c	f_i	f_c	Δ	δ	*Rank*
Immediate family	*10094*	*99*	*12.3*	*6.2*	*-6.0*	*-49.1*	*10*
Extended family	*5104*	*49*	*6.2*	*3.1*	*-3.1*	*-50.2*	*11*
Internet	*5271*	*54*	*6.4*	*3.4*	*-3.0*	*-46.8*	*9*
Grew up together	*3253*	*26*	*4.0*	*1.6*	*-2.3*	*-58.5*	*13*
Friend of family	*3820*	*47*	*4.6*	*3.0*	*-1.7*	*-36.1*	*6*
Live(d) in same neighborhood	*2483*	*26*	*3.0*	*1.6*	*-1.4*	*-45.7*	*8*
Travel/Exhange/Pen pal	*1867*	*16*	*2.3*	*1.0*	*-1.3*	*-55.5*	*12*
Mutual friend	*7656*	*130*	*9.3*	*8.2*	*-1.1*	*-11.9*	*3*
Hobby/Sport/Interest	*3365*	*50*	*4.1*	*3.2*	*-0.9*	*-22.9*	*5*
Other	*913*	*11*	*1.1*	*0.7*	*-0.4*	*-37.5*	*7*
Faith/Volunteering	*1474*	*22*	*1.8*	*1.4*	*-0.4*	*-22.5*	*4*
School	*17333*	*410*	*21.1*	*25.9*	*+4.8*	*+22.8*	*2*
Work	*19628*	*645*	*23.9*	*40.7*	*+16.8*	*+70.5*	*1*

Table 2.3 Comparisons of the initiation of relationships in complete and incomplete chains. Categories are ordered according to increasing Δ; all quantities are defined the same as in the previous analysis (see Table 2.2).

Strength	N_i	N_c	f_i	f_c	Δ	δ	*Rank*
Extremely close	*15268*	*145*	*18.6*	*9.1*	*-9.4*	*-50.7*	*5*
Very close	*18848*	*225*	*22.9*	*14.2*	*-8.7*	*-38.0*	*4*
Fairly close	*26824*	*473*	*32.6*	*29.8*	*-2.8*	*-8.5*	*3*
Not close	*3319*	*140*	*4.0*	*8.8*	*+4.8*	*+118.9*	*1*
Casually	*18000*	*602*	*21.9*	*38.0*	*+16.1*	*+73.6*	*2*

Table 2.4 Comparisons of the strength of relationships in complete and incomplete chains. Categories are ordered according to increasing Δ; all quantities are as defined in the previous analysis (see Table 2.2).

Reason for choosing link	N_i	N_c	f_i	f_c	Δ	δ	Rank
Geography	26002	237	35.9	21.8	-14.0	-39.1	6
Travel	10455	48	14.4	4.4	-10.0	-69.3	7
Continue the chain	5563	7	7.7	0.6	-7.0	-91.6	9
Lots of friends	5493	19	7.6	1.7	-5.8	-76.9	8
Family origin	7818	74	10.8	6.8	-4.0	-36.8	5
Work	1616	48	2.2	4.4	+2.2	+98.3	3
Similar education	2770	85	3.8	7.8	+4.0	+104.9	2
Other	6664	172	9.2	15.8	+6.6	+72.3	4
Similar profession	6130	396	8.5	36.5	+28.0	+331.3	1

Table 2.5 Comparisons of reasons given by participants in complete and incomplete chains for choosing next individual. Categories are ordered according to increasing Δ; all quantities are as defined in the previous analysis (see Table 2.2).

With respect to the individual characteristics, there was a strong trend that participants with higher income more often appeared in successful chains. Although young people (18-29) dominated the age range for participants, complete chains showed high preponderance for participants within the medium age range (30-39). Males and those with graduate education disproportionately accounted for successful chains (Table 2.6).

Income	N_i	N_c	f_i	f_c	Δ	δ	Rank
Low	14046	64	22.9	18.7	-4.2	-18.5	4
Very low	3512	9	5.7	2.6	-3.1	-54.2	5
Average	21957	122	35.8	35.6	-0.2	-0.6	3
High	15510	93	25.3	27.1	+1.8	+7.2	2
Very high	6306	55	10.3	16.0	+5.8	+56.0	1
Age							
18-29	32210	155	39.7	34.3	-5.4	-13.5	4
Above 60	3644	9	4.5	2.0	-2.5	-55.6	6
17 and under	454	2	0.6	0.4	-0.1	-20.8	5
50-59	9389	52	11.6	11.5	-0.1	-0.5	3
40-49	12469	70	15.4	15.5	+0.1	+0.9	2
30-39	23060	164	28.4	36.3	+7.9	+27.8	1
Education							
High school	11460	19	15.4	4.4	-11.0	-71.3	4
College	39097	185	52.6	43.1	-9.5	-18.0	2
E. school	654	2	0.9	0.5	-0.4	-47.0	3
G. school	23123	223	31.3	52.0	+20.9	+67.1	1
Gender							
Female	51110	265	56.9	54.1	-2.8	-5.0	2
Male	38670	225	43.1	45.9	+2.8	+6.6	1

Table 2.6 Comparisons of the demographics of participants in complete and incomplete chains. Categories are ordered according to increasing Δ; all quantities are defined as in the previous analysis (see Table 2.2).

We also asked participants to state the reason why they chose a particular individual as the next person in the chain (Figure 2.3). Two factors—geographical

and occupational proximities—stood out as the most-used cues for directing messages. Specifically, if we saw the reason chosen as a function of the chain length, as displayed in Figure 2.3, in the early stages of the chain, it appeared that geographical reason dominated, presumably because senders were geographically distant from targets. Yet, at the later stages of the chain, occupational cue was used more than geographical cue. This finding suggested that when proximity to a target in a domain (e.g., geography) has reached a certain level of granularity at which it was too difficult to go further, senders switched to another domain (e.g., occupation) that could provide further differentiation. These results illustrated the importance of cross-cutting social domains and switching across these domains (White 1992) to the ability of individuals to navigate social networks.

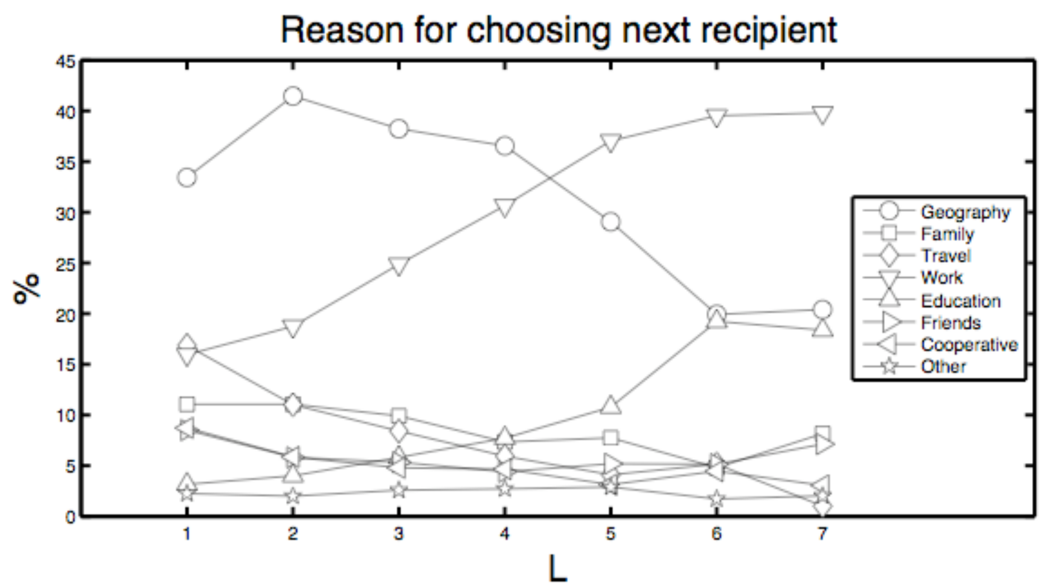

Figure 2.3 Reasons for choosing the next recipient. *L* is the number of steps in chains. Geography, recipient is geographically closer; Family, recipient's family originates from target's region; Travel, recipient has traveled to target's region; Work, recipient has occupation similar to target; Education, recipient has similar educational background to target; Friends, recipient has many friends; Cooperative, recipient is considered likely to continue the chain.

We found that senders rarely chose an acquaintance because he or she had many friends. In fact, as shown in Table 2.5, participants in successful chains were far less likely than those in incomplete chains to send messages to hubs (1.7 versus 7.6%). Thus, at least within the context of our experiment, the presence of highly-connected individuals had little importance. In addition, in contrast to Milgram's experiment, we did not observe the "funneling" effect where a single individual was responsible for the majority of messages that were received by a target. Figure 2.4 shows the percentage of messages that reached targets through senders who sent one message, two messages, and so on. At most, 2% of messages passed through a single acquaintance of any target, and 85% of all chains reached targets through individuals who delivered at most three messages. Ostensibly the search process in the small-world experiment did not require exceptional individuals who were disproportionately responsible for completed chains; instead, it could be done in an egalitarian fashion.

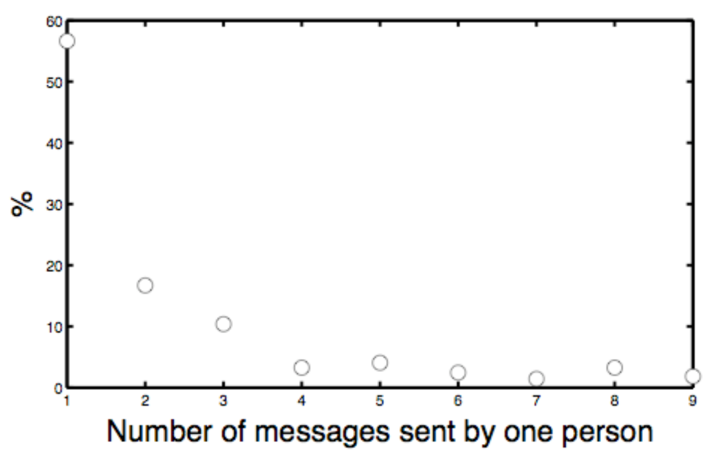

Figure 2.4 Percentage of messages sent by senders who sent *x* number of messages. Almost 60% of messages reached targets through a unique individual. At most, 2% of messages reached a target through a single person.

We observed a tendency for individuals to send messages to someone similar to themselves (Figure 2.5) (Lazarsfeld and Merton 1954; McPherson, Smith-Lovin and Cook 2001). For example, there was a clear tendency to send messages within the same age group and gender.[7] The dominant trend with respect to income and educational variables also showed the pattern of homophily. However, there were exceptions where people with the lowest education levels were more likely to send messages to higher-educated individuals. One possible explanation was that the targets' educational and income levels were relatively high, and lower-status participants tended to send messages to "higher-status" people than themselves. In other words, messages tended to move to people similar to the targets.

[7] In the case of gender, it is interesting to note that Travers and Milgram reported a similar tendency but with much stronger effect; men were ten times more likely to send messages to other men than to women, raising the speculative possibility that social changes in the last forty years have decreased gender barriers substantially.

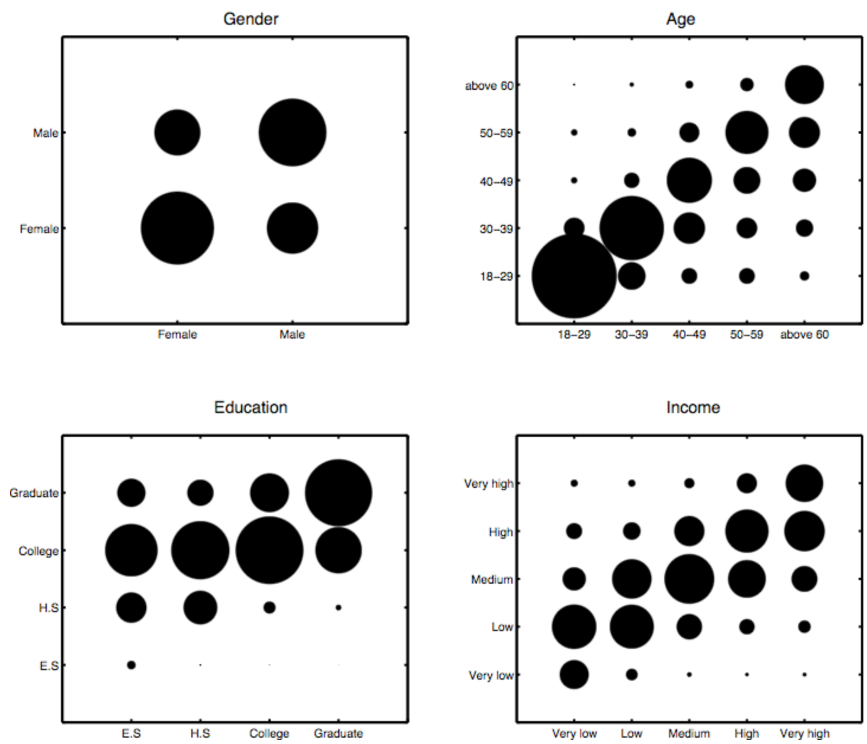

Figure 2.5 Links comparison based on gender, age, education, and income. Horizontal axes are sender's attributes and vertical axes are recipient's attributes. Area of the circles is proportional to the percentage of messages sent between the corresponding categories.

In general, our data suggests that the progress of messages did not follow a hierarchical pattern where people with higher or lower socio-economic status were preferred. Instead, links tended to be localized in social space, where similar people were more likely to send messages to each other. This could be the effect of homophily, in which social distance was mapped into network distance (McPherson, Smith-Lovin and Cook 2001). Because of homophily, we tend to be surrounded by people with similar backgrounds and interests. Others who came from different social backgrounds, or were involved in unfamiliar activities, seemed to be very far away socially. Yet, distance in social space was not always correlated with distance in network space. Because of homophily and

our limited intuition, people tended to conflate social and network distances and assume that those with very different social characteristics are also very far in terms of network distance. This was not necessarily the case.

People only have local information regarding their personal network and thus, it is very likely that people do not always make the best choice for the shortest path (Killworth et al. 2005). The fact that some chains still reached their targets showed that the aggregation of individuals with limited knowledge can still complete the task of finding short paths. There is no need for initial senders to completely preconceive the chain that will eventually link them to target persons. The task needs to be solved collectively, and our results suggest that people can actually achieve it. The small-world experiment shows that agents with local information and limited reasoning power can solve ambiguous problems collectively; a problem that seems impossible on a local scale is resolved on a higher scale.

One striking result of the experiment is the completion rate, which is very low; only 491 chains reached their targets, which is a 0.5% completion rate. The completion rates for our experiments, however, are much lower than those recorded by Milgram and his colleagues: In experiment 1, a total of 491 chains (0.5%) successfully reached their targets. Out of eighteen targets, only one target, a professor in New York, obtained a completion rate of more than one percent, and eight targets never received any messages. The reason for these ultra-low completion rates is easily traced to attrition rates in our experiments that were considerably higher than those experienced by Milgram—25% versus 67%

for experiment 1. The peculiar design of the small-world method, moreover, causes chain completion rates to diminish exponentially with chain length. For example, if one hundred thousand chains are initiated with a 25% constant attrition, after six removals there are 1780 chains left, whereas with a 67% attrition rate, only 129 survive. Therefore, completion rates are highly sensitive to attrition rates and hence we need to have some ideas about the reason behind chain termination to be able to interpret the low completion rates properly.

There are three possibilities of why attrition occurred. First, attrition could occur randomly, because of apathy, technical difficulties or refusal to participate in the experiment. Attrition could also occur disproportionately at longer chain lengths, which means that the chains get "lost" or are otherwise unable to reach their targets. Another possibility is that attrition occurs at short chain lengths, because individuals far away from the target are less likely to pass the message. Our results, depicted in Figure 2.6, support the random-failure hypothesis because the attrition rate remains almost constant for all chain lengths except for the first step, which is a special case because senders registered voluntarily rather than receiving a message from an acquaintance.

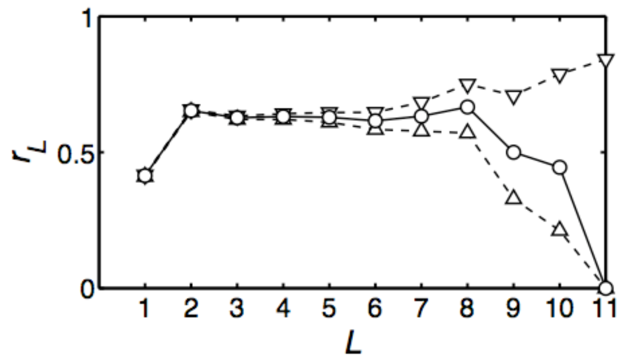

Figure 2.6 Average per-step attrition rate (circles) and 95% confidence interval (triangles).

Since attrition is mainly the result of unreliability in the measurement device used to probe the connectivity of social networks, then it is appropriate to take an active stance toward the data[8] (Leifer 1992) and investigate the condition of an “ideal” world, free from measurement error. For Milgram’s data, the distribution of chain lengths in the hypothetical condition of a 100% response rate has been calculated (White 1970).

[8]If we drop a stone and a leaf together in a non-laboratory condition, because of friction with air, the leaf will touch the ground later. To observe the pure effect of gravity we need to perform the experiment in a vacuum, where both the stone and leaf will hit the ground at the same time. Leifer (1992) pointed out that mature science requires an active approach to the data, in which observers create a working environment in which they can see the effect predicted by theory.

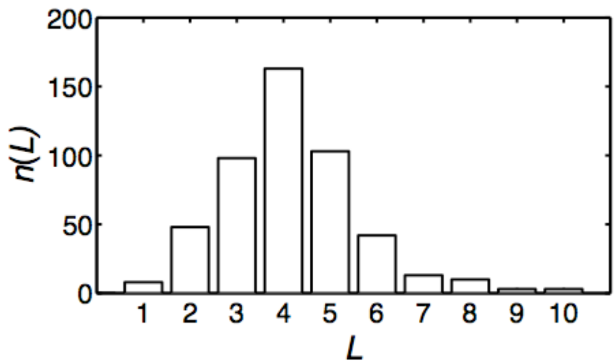

Figure 2.7 Histogram representing the number of chains that are completed in L steps ($\langle L=4.2\rangle$).

Applying this active stance to our data, we then ask what the distribution of completed chains would look like without this stochastic attrition? From our distribution of 491 completed chains (Figure 2.7), the average number of steps was $\langle L\rangle = 4.21$. This result, however, is misleading. The longer a chain is, the greater the chance the chain will result in failure because of stochasticity. Thus, shorter chains are more likely to reach their targets. Hence, the sample from which we take the average is made up of relatively short completed chains and hence it is biased.

We will develop a rigorous method to calculate the estimators of the chain distribution in the next chapter. For the remainder of this chapter, we will discuss a heuristic to do the estimation that can provide the intuition behind the more formal derivation. As a matter of fact, the result from this intuitive derivation turns out to be unbiased, as we will see in Chapter 3. In addition, the resulting estimator turns out to be equivalent to the original estimator used by White (1970) with one important exception on the probability of completing the final

links if the senders know the target. White assumed that senders who know the target always send the message. We relax this assumption, because we have evidence that the last links were not always completed.

To start, imagine an “ideal” world where people follow total compliance in any social science survey, and hence produce a 100% participation rate. In this ideal world, with zero attrition rates, a message either reaches its target or continues to the next person in the chain (Figure 2.8A). In an “ideal” world, for each step L, there are only two variables: the number of incomplete chains (i.e., active chains that have not reached their targets yet, $\hat{N}_{I,L}$) and the number of completed chains (i.e., reached their targets, $\hat{N}_{C,L}$). In the real world, people drop out from participation for reasons ranging from individual apathy, reluctance and lack of time, to technical computer problems. Thus at each step, failures to continue the chains produce attrition (Figure 2.8B). Thus, in the real world, there are incomplete ($N_{I,L}$) chains, completed ($N_{C,L}$) chains, and because of attrition (r_L), there are also failed ($N_{F,L}$) chains.

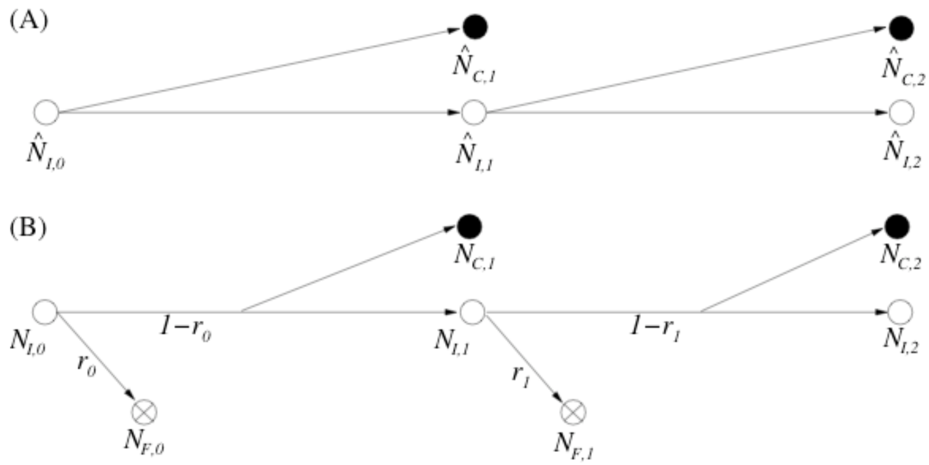

Figure 2.8 Illustration of the progression of chains with and without attrition (A). In a hypothetical world with no attrition, participants always pass on messages. From the initial number of senders ($\hat{N}_{I,0}$), messages either reach targets ($\hat{N}_{C,1}$) or continue to the next step, where $\hat{N}_{I,1} = \hat{N}_{I,0} - \hat{N}_{C,1}$ is the number of messages at step one. For the second step, some messages reach targets ($\hat{N}_{C,2}$) or continue as incomplete chains ($\hat{N}_{I,2}$), and so on (B). In the real world, some of initial senders ($N_{I,0}$) do not pass messages, with the probability of r_0, so there are failed messages at step zero ($N_{F,0}$). The rest of the messages $(1-r_0)$ are either completed ($N_{C,1}$) or stay incomplete ($N_{I,1}$) at step one. This process continues for the next step. The cumulative effect of attrition renders an exponential decrease in the total number of messages at each step.

Referring to Figure 2.7A in an ideal world, all starting messages ($L = 0$) are transmitted to the next step ($L = 1$), and they either reach the target or not, so we can write

$$\hat{N}_{I,0} = \hat{N}_{I,1} + \hat{N}_{C,1}. \tag{2.1}$$

In the real world, however, only a fraction of initial messages are forwarded to the next step, so we obtain

$$(1-r_0)N_{I,0} = N_{I,1} + N_{C,1}. \tag{2.2}$$

Importantly, the initial number of messages is the same for both the ideal and the real world, that is

$$\hat{N}_{I,0} = N_{I,0} \,. \tag{2.3}$$

Thus, combining the three equations above we get

$$\hat{N}_{I,1} + \hat{N}_{C,1} = \frac{N_{I,1} + N_{C,1}}{(1 - r_0)} \,. \tag{2.4}$$

Invoking the assumption that the attrition rate is unrelated to the underlying network structure and the search process, we can separate each term for completed and incomplete chains because they come from similar populations and, for each of them, obtain the expression that relates observed and predicted distribution without attrition. For completed chains, in the first step we obtain

$$\hat{N}_{C,1} = \frac{N_{C,1}}{(1 - r_0)} \,. \tag{2.5}$$

Using the same procedure, we repeat the calculation for $L = 1$ and obtain

$$\hat{N}_{C,2} = \frac{N_{C,2}}{(1 - r_0)(1 - r_1)} \,. \tag{2.6}$$

More generally, for the number of completed chains with length L when there is no attrition, $\hat{N}_{C,L}$ is given by

$$\hat{N}_{C,L} = \frac{N_{C,L}}{\prod_{j=0}^{L-1} (1 - r_j)} \,, \tag{2.7}$$

where $N_{C,L}$ is the observed number of completed chains in L steps and r_L is the probability of attrition from step L to $L + 1$.

Using equation 7, we are able to produce an estimate of an ideal completed chain distribution. In specifying the distribution, however, we must calculate the median instead of the mean. Because there are only a few chains with a length of seven steps or more, the estimation of chain length at the tail of the distribution (long chains) gives high variability. If we take the mean of this distribution, then we will grossly overestimate the average chain length. Therefore, we obtain the median of the distribution $L_{med} = 7$. Hence, if we hypothetically assume an ideal world with zero attrition, then the typical path length connecting two individuals is seven. [9]

The preceding estimation procedure, however, uses only completed chains and ignores most data points, which are incomplete chains, and so it produces biased estimates. Moreover, the calculation assumes that everyone has the same attrition. In the next chapter we will show that the homogeneous attrition assumption is unattainable, so we need other methods to estimate the chain length distribution that can include attrition heterogeneity. As it turns out, a novel statistic method can solve both problems simultaneously: giving unbiased estimates of chain length distribution and using heterogeneous attrition assumption.

---

[9] In our paper Dodds, Peter S., Roby Muhamad, and Duncan J. Watts (2003), we decomposed the distribution according to whether initial senders and targets reside in the same country or not. The typical path length for chains that started and ended in the same country is five, and the median path length is seven if the chains include cross-country connections. Therefore, the estimated range of the typical median path length is between five and seven, depending on the geographical separation between the sources and targets.

CHAPTER THREE: THE ATTRITION PROBLEM[10]

One recurring finding from large-scale small-world experiments was that only small proportions of chains successfully reached their ultimate targets, and these successful chains were typically short. For example, in our own experiment there were only 491 (0.5%) completed chains with the average length of about four, while the rest of the 105,804 chains never reached their targets. Travers and Milgram obtained 6% and 30% completion rates with the average chain length of eight and six for the Kansas and Nebraska studies respectively (Travers and Milgram 1969). Korte and Milgram tried to connect white senders in Los Angeles to black targets in New York, and achieved a 13% completion rate with the average length of seven for completed chains (Korte and Milgram 1970). Although the vast majority of chains in small-world experiments never reach their ultimate targets, most of the attention has been on those few short completed chains.

Thus, most conclusions from small-world experiments regarding the distribution of chain length is based only on small portions of chain data, while most of the chain data is actually missing.

In the previous chapter, we constructed an estimator using a heuristic on chain progressions. This informal approach suffered from two shortcomings: (1) it requires the assumption of homogeneous attrition, and (2) we did not account for the possible bias in the estimator. We will address these two problems in the

---

[10]Some of the results in this chapter have been reported in Goel, Sharad, Roby Muhamad, and Duncan J. Watts. 2009. "Social Search in "Small-World" Experiments." *WWW 2009*.

current chapter. First, we will focus our attention toward understanding the characteristics of attrition in small-world experiments that are responsible for the large amount of missing data. Then, we will construct a statistical method that can take into account the missing data and hence produce an unbiased estimate for the distribution of chain length. Finally, using the new unbiased estimator, we reconstruct chain length distributions for both cases of homogeneous and heterogeneous attrition rates.

3.1 Stochastic attrition and its critique

In the previous chapter we have shown that results from our experiment lend support to the idea that attrition in the small-world experiment was stochastic, i.e., the drop-off in participation is neither related to the underlying network structure nor the search process, instead it is related to insufficient motivation to pass on messages or failure to receive the message in the first place. Travers and Milgram also used this stochastic interpretation of attrition in their original paper, where they found no statistically significant differences between people in complete and incomplete chains[11]. In addition to the absence of evidence to think otherwise, the stochastic attrition assumption also allows us to estimate the “true” length distribution of chains that would have been observed had no attrition taken place (White 1970). The average chain length if attrition is not present is longer than the observed distribution because attrition renders longer chains to be more likely to terminate (see Figure 3.1). However, as we

[11] Travers and Milgram collected data on dropouts by asking each sender the age and gender of the recipient, including the nature of their relationship and the reason why the particular recipient was selected.

have seen in the previous chapter, although the “true” chain length is longer than the observed chain length, the resulting estimates of the true chain are still “short.” Thus, as long as the stochastic failure assumption is reasonable, the conclusion from most small-world studies, even with low completion rates, is sufficient to support the small-world hypothesis.

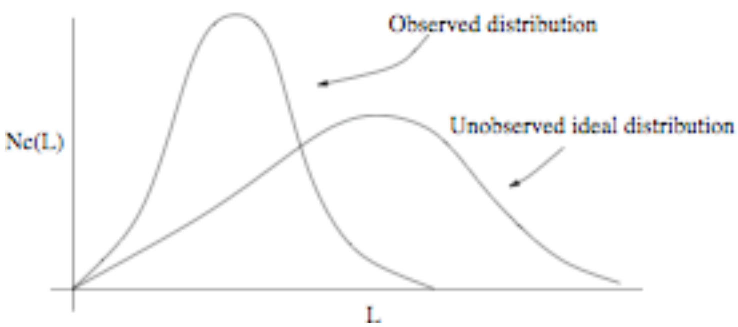

Figure 3.1 The assumption that the message-passing process is a stochastic one implies that the observed completed chains’ length distribution is a modification of an ideal distribution in which attrition is not present. Due to attrition, there is a bias toward shorter chain length for the observed distribution.

There is, however, an alternative interpretation that asserts that attrition is not stochastic, but it could be related to the topology of social networks and the search process itself; in other words, attrition could indicate that most people are separated by chains that are too long and hence cannot be completed. For example, Kleinfeld (2002) is one of the proponents of this alternative interpretation and she wrote:

> The research on the small-world problem suggests not a counter-intuitive triumph of social research, but an all-too familiar pattern: We live in a world where social capital, the ability to make personal connections, is not wide-spread and more apt to be the possession of high-income, white people or people with exceptional social intelligence. Certainly some people operate in small worlds, such as scientists with worldwide connections or university administrators, but many low-income people or

> minority people do not seem to. What the empirical evidence suggests is that some people are well-connected and others are not, a world not of elegant mathematical patterns where a random connector can zap us together but a more prosaic world, a lot like a bowl of lumpy oatmeal, with many small worlds loosely connected and perhaps some small worlds not connected at all.

The critique by Kleinfeld can actually be separated into two distinct but related aspects (Goel, Muhamad and Watts 2009). The first aspect is related to social capital. According to this argument only certain people posses the ability and the resource to conduct searches in their social networks effectively; presumably, these are individuals with high social status or strong social skill. Thus, the differentiating factor here is individual heterogeneity in terms of social capital. Here, incomplete chains indicate not so much of isolation because of the lack of connection per se, but more because of the lack of "well-connected" individuals who have the unique capability to mobilize their social capitals. Accordingly, this interpretation leads us to the possibility that attrition is a function of individual attributes such as gender, education, and age, known to be correlated with social capital. If this is true, then attrition cannot be assumed as mere stochastic fluctuation that is not related to the search process, but attrition that reflects individual-level heterogeneity.

The second aspect of the critique is the possibility that there are individuals who reside in separate populations who can only be connected through long or otherwise undiscoverable paths (Figure 3.2). Individuals who live within the same population, e.g., sharing geographic and demographic attributes, are connected with "short" paths, but since most people are "far away" from each

other, thus—at least in small-world experiments—only a small portion of chains can be completed. This second objection implies that because most chains are long, then the estimation of the distribution of chain length must take into account long chains if we want the estimator to be unbiased. Consequently, the estimator used so far to estimate the ideal chain length distribution (equation 2.7) is problematic because it uses the assumption that attrition is homogeneous *and* it does not take into account long chains, and hence there is no guarantee the estimator is unbiased. These two aspects of attrition, the lack of social capital and short paths, can combine and aggravate the attrition problem.

In summary, there remain two unanswered questions in the current small-world literature. The first question is about the characteristic of attrition, especially regarding whether a stochastic process generated attrition and hence resulted in attrition that was uniform across individual attributes. The second question is how to produce an unbiased estimate of chain length distribution that incorporates long chains that are missing from data. The rest of this chapter is dedicated to answer these two questions.

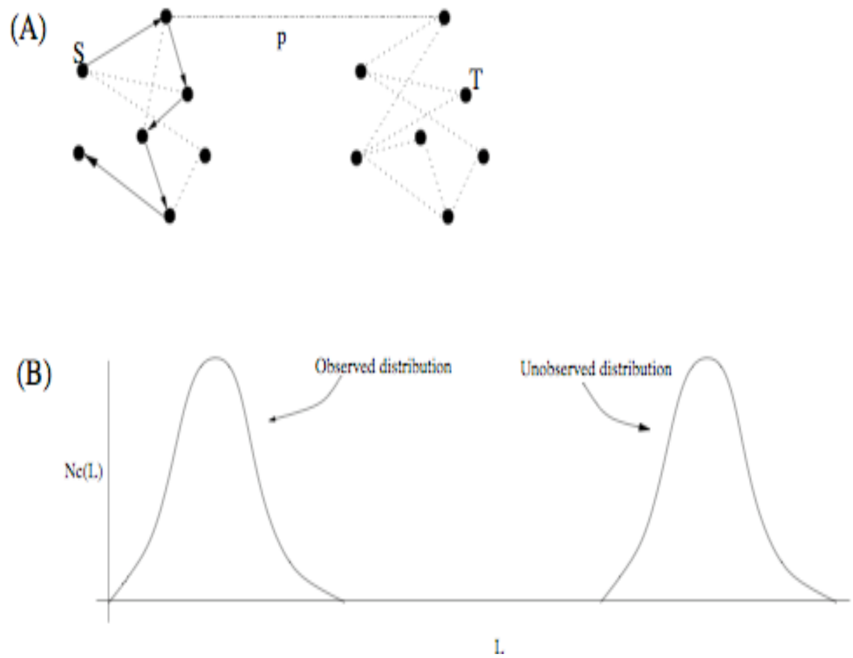

Figure 3.2 (A) In the stochastic interpretation, the process of passing messages from a source *S* to a target *T* includes a probability of failure. Thus, the termination of message chains does not imply the absence of an underlying path; here the chains do not complete even though there are connections *p* that could be very long. (B) The assumption that the absence of connection gives rise to attrition implies a bimodal distribution. The observed distribution represents all chains that can be completed; these are chains in a homogenous population. The second mode is the unobserved distribution of very long, possibly infinite chains that cannot be completed.

3.2. The characteristics of chain attrition

Before we perform a rigorous analysis of attrition, we begin by conducting a descriptive analysis of attrition using data from two versions of our experiments. The first of these experiments was implemented between December 2001 and August 2003; the second version followed immediately thereafter, and ran until December 2007. In the first experiment, 98,865 people from 168 countries initiated 106,295 chains directed at eighteen targets in thirteen countries. In the second experiment, 85,621 people from 163 countries participated in 56,033 chains, directed at 21 targets in thirteen countries. In both

experiments, most participants were from North America and Western Europe, white, Christian, and predominately young, college-educated, middle-class professionals.

From Table 3.1, we can observe that our results exhibit the same combination of short path lengths and low completion rates that typify small-world experiments. To understand better the origins of attrition in small-world experiments, we begin with the simplest available analysis—namely, comparing the average attrition between the first and second versions of the experiment. We emphasize that both versions exhibited the same basic design, and that in fact the second version incorporated a number of design improvements over the first, including an improved user interface, a more detailed set of survey questions, more comprehensive target descriptions, and the option for participants to forward messages to more than one friend. If the primary basis for chain attrition is the inability of individuals to locate suitable contacts, one would expect either that both experiments would display similar attrition, or that attrition would be lower for the second experiment (the latter, in fact, was precisely the intention of the improvements).

In contrast to expectation, however, attrition in the second version was considerably higher than in the first—even for targets that participated in both versions—and completion rates correspondingly dropped to 20% of their previous value. What happened? Clearly the world did not somehow become less connected somewhere around August 2003, nor did social search become more difficult. What did change during the period between the start dates for the

experiments, however, was that the incidence of junk e-mail or “spam” increased by roughly 1000% (MWAAW 2006)—a trend that continued for the duration of the second experiment (in 2006, it was estimated that 80% of all e-mails were spam). Unsurprisingly, along with this exponential growth in the rate of junk mail was a corresponding improvement in “spam filters”; thus, it is extremely likely simply that many of our messages never reached their intended recipients[12]—a conclusion that is supported by a number of anecdotal reports sent to user-support e-mail.

	City	Country	Occupation	Sex	$r(r_0)$	N	$N_c(\%)$	<L>
1	Bronx	USA	Exec. Chef	M	.81(.42)	2435	2(.1)	4
2	Santiago	Peru	Director	M	.71(.42)	2902	1(.03)	6
3	Gainesboro	USA	Welder	F	.64(.44)	2311	0	n/a
4	Sydney	Australia	Homemaker	F	.75(.40)	2562	0	n/a
5	Sau Paulo	Brazil	Entrepreneur	F	.73(.40)	2889	6(.2)	4.3
6	Para-aque	Philippines	Teacher	M	.70(.45)	2710	0	n/a
7	Grand Island	USA	Waitress	F	.67(.49)	4019	0	n/a
8	Melbourne	Australia	Air Traffic Controller	F	.69(.49)	2511	0	n/a
9	Singapore	Singapore	Teacher	F	.63(.40)	3059	4(.1)	1
10	Cape Town	South Africa	Astrologer	F	.76(.44)	2643	2(.08)	1.5
11	Salem	USA	Mother	F	.72(.41)	2831	3(.1)	4
12	Kingstown	St.Vincent	Nurse	F	.44(.43)	2932	0	n/a

Table 3.1 List of targets for experiment 2. Average and initial attrition rates are denoted by r and r_0 respectively. N is the number of chains assigned to the corresponding target, N_c is the number of chains that reached targets, and <L> is the mean path length of completed chains.

[12] Although participants were forwarding e-mails to individuals whom they knew, for practical purposes it was necessary to have them perform this task using our web server. Thus, although the next recipient was, in fact, receiving an e-mail request from a trusted friend, it would have appeared to their mail server that it was coming from an unknown address.

Next, we examine the stages of the experimental procedure in which the most attrition has occurred. In particular, we are interested to know whether most attrition occured before or after participants see target information. To answer this question, we used data from the second version, because in that version, information about targets was not included in the invitation e-mails. Recipients received only minimal information about the experiment and were given a link that would take them to the experiment website. Figure 3.3 depicts the percentage of participants who did not complete some stage of the experimental procedure. Brief information about targets was displayed between stages 3 and 4. Full information about targets was available after completing the survey (between stages 4 and 5).

In total, 80.6% of participants failed to continue their chains. Most of this attrition occurred before they came to the website, so 56% of messages were terminated before the intended recipients knew who their targets were. Only about 6% of messages were terminated after the senders were informed about their targets. We can assume that participation includes two decisions: the decision to participate and the decision to choose the next recipient. This data suggests that most attrition happened in the first decision. The decision to participate is arguably not affected by network connectivity or the perception thereof, because no information about targets was available at that step. Therefore, our data suggests that non-structural factors, such as individual apathy or filtering technology that mistook our messages as junk, contributes to the majority of attrition in this experiment.

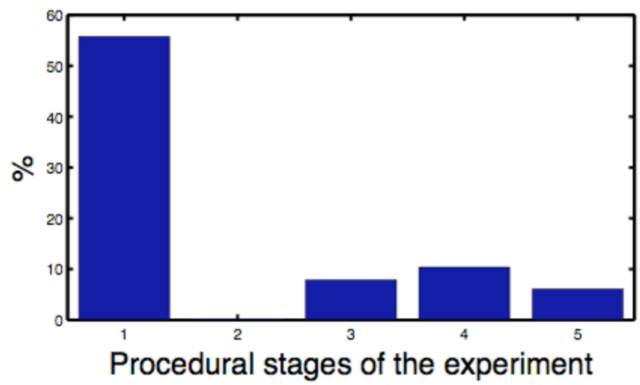

Figure 3.3 Bars represent the percentage of participants who did not complete the corresponding procedural step in the experiment. Participants saw brief information of targets for the first time between stages 3 and 4, then saw full information of targets between stages 4 and 5. These stages are: 1=Visiting the website, 2=Verifying the sender, 3=Registration, 4=Demographic survey, 5=Relationship survey and selecting the next recipient.

A third way to examine the factors causing attrition is to consider the geographical progression of chains. Here, we used data from the first version of the experiment and focused on chains that originated from outside the countries where the targets were located. We examined at which geographical stages most attrition occurred: whether when trying to reach the right country, the right city, or the target itself. We found that most of the attrition occurred when participants were trying to reach the right country (Figure 3.4). It is implausible to ascribe the difficulty of reaching the targets' countries to the absence of connections, simply because it is hard to think of any country that is completely isolated. For example, consider Target #10 in the United Kingdom. More than 90% of messages could not reach that country. If we attributed the attrition to the lack of

connection, then we would have to conclude that the United Kingdom is an isolated country. This purely structural interpretation obviously cannot be applied here. Therefore, high attrition alone cannot be used as an argument for non-connectivity.

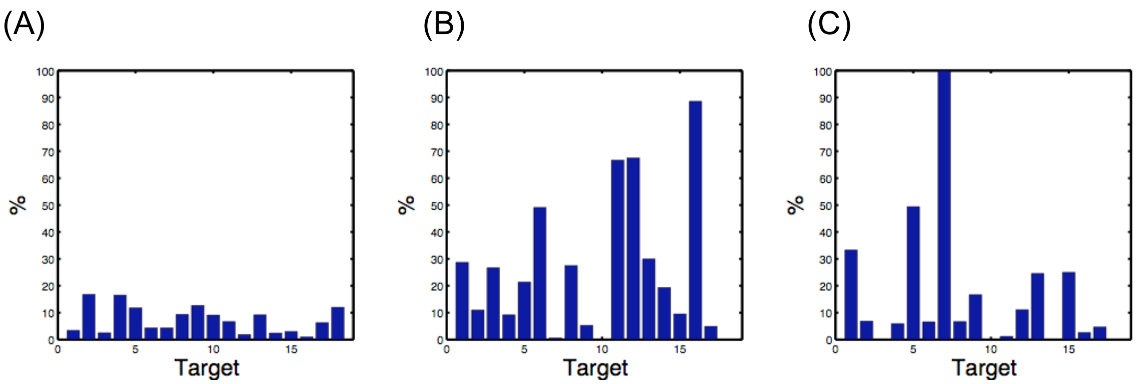

Figure 3.4 (A) Bar represents the percentage of chains originating outside targets' countries and reaching the right country. (B) The percentage of chains that reached the right cities after getting into the right country. (C) The percentage of chains that reached the targets from the right city. Most attrition occurred when trying to get into the right countries. Since there is no country that is completely isolated, it is problematic to think that this attrition was caused by the lack of connections.

3.3 Modeling attrition

The discussion from the previous section has shed some light on some characteristics of attrition in the small-world experiment, but it did not directly address the problem of attrition heterogeneity, because to do so we need to explicitly model attrition in terms of individual attributes that contribute to the variation in social capital. There is, however, a problem in modeling attrition, because participants who did not continue messages never came to our website in the first place and so we do not have data about their demographic

characteristics; thus, we cannot directly estimate neither the probability of continuance nor the probability of dropoff. Thus, as a proxy, we instead estimate the probability of “next-step continuance” which is defined as the following: For every pair of sender (A) and receiver (B) where the receiver is not a target, we estimate the probability for the receiver to continue the chain based on A’s individual attributes and relational attributes between A and B. This probability of continuance can be interpreted as a measurement of search ability of participants, i.e., the ability of A to pick someone who will continue the chain. In total, we analyzed 88,875 links, of which 32% of them are continued links (the recipient forwarded the message) and 68% of the links were terminated (the recipient did not continue the chain).

We model the next-step continuance probability by logistic multilevel regression, which is suitable for data with group structure. In our data, groups are levels within a category; for example, for the education category, there are separate groups for “elementary school,” “high school,” “college,” and “graduate school.” Multilevel regression is a middle ground between two extreme approaches for modeling data with group structures. One extreme is treating different groups within the same category as unrelated to each other (no pooling), so each group has its fixed parameter with no relationship with each other; for example, within the education category, there is a fixed parameter for each level of education. The other extreme ignores the groups within a particular category (complete pooling); for example, the complete pooling approach for the education category treats all individuals the same regardless of their education

levels. In other words, whereas no pooling could overestimate the variation between groups, especially if the sample size within groups are small, complete pooling ignores variation between groups. Multilevel models can be seen as partial pooling that lies in the middle ground between these two extremes approaches; partial pooling allows the possibility that groups within a category are related but without imposing a hard constraint on the strength of their relationship.

Our multilevel model can be written as:

$$P(y_i = 1) = \text{logit}^{-1}\left(\gamma + \beta_{\text{nonwhite}} X_{\text{nonwhite},i} + \beta_{\text{female}} X_{\text{female},i} + \sum_{k=1}^{9} \alpha_{j_k[i]}\right)$$

where the outcome variable y_i indicates the next-step continuance, γ is the intercept, the two β terms are fixed effects for non-white and female participants respectively, and the $\alpha_{j_k[i]}$ corresponds to the nine group effects. For each category k (e.g., education), $j_k[i]$ is the group (e.g., high school, college) of the i^{th} link; we model the group parameters within each category as coming from a normal distribution: $\alpha_{j_k[i]} \sim N(0, \alpha_k^2)$. In total there are 66 variables: a common intercept variable; one variable each for gender (male/female) and race (white/non-white); 54 attribute variables that are grouped into nine categories[13] (age, education, work field, work position, income, strength of relationship, reason for choosing recipient, origin of relationship, and target); and one variance parameter for each category σ_k^2.

[13]We found that religion, country, type of relationship, and current position in the chain were not statistically significant, so we excluded them from the final model.

When σ_k^2 is small, it indicates that there is little variation (strong association) among groups within the corresponding category k, and when the variance is large, it indicates weak association among groups within category k. The standard deviations σ_k for the nine categories are depicted in Table 3.1, sorted from the highest to the lowest. To make interpretation easier, the attrition column shows the result of transforming the standard deviation to a probability scale that is relative to the baseline attrition rates. Thus, for example, differences between age groups account for a 2% absolute change in attrition rates. The education category has the highest standard deviation, so the differences in education levels contribute about 3% to absolute changes in attrition rates. Since the baseline attrition is 30% for white males, then 3% absolute difference corresponds to 10% relative difference. At first glance, the difference is small but, as we will discuss below, the difference is amplified by the correlation among attributes and also compounding effect as chains propagate.

Category	σ_k	Attrition
Education level	0.14	± 0.03
Age	0.12	± 0.02
Relationship strength	0.11	± 0.02
Target	0.09	± 0.02
Work field	0.08	± 0.02
Income	0.07	± 0.01
Reason for choosing recipient	0.05	± 0.01
Relationship origin	0.04	± 0.01
Work position	0.03	± 0.01

Table 3.2 Category standard deviation parameters from a multilevel logistic regression model for next-step continuance probabilities. Attrition is presented as typical deviation from the baseline of 30% for white males.

Now we turn to the analysis of the effects of groups within each category. Table 3.2 depicts the analysis for individual attributes on the next-step continuance probability relative to the baseline probability of 30% for typical white males. Tables 3.3 and 3.4 provide the same analysis from relational attributes and targets respectively. Each row in each table corresponds to a group (e.g., "18-29" age group, "Graduate school") within attribute categories (e.g., "Age," "Education level") and the overall intercept and two fixed effects for females and non-whites. For any given group, the second column is the estimated regression coefficients for that group in the logarithmic scale along with its associated standard error, and the second column is the corresponding effect on the probability of next-step continuance relative to the baseline probability.

Consistent with Table 3.2, Tables 3.3, 3.4, and 3.5 reveal a small but significant range of attrition rates. In terms of individual attributes, having a graduate degree has the highest effect by increasing the probability of continuing

a chain by 4%; relatively young and high-income participants are also better in passing along messages by 3% and 2% respectively. In contrast, having only a high-school degree diminished the probability of passing along messages by 3%, and participants with low income were less likely than average to continue chains by 1%. Comparisons of typical white males, females and nonwhites show smaller probability of next-step continuance by 1% and 3% respectively. Overall, participants with high socio-economic status are more likely than average to continue chains, and hence their searches are more likely to be successful.

In terms of relational attributes, the strength of relationships has the largest effect. Specifically, if messages were passed via the strongest link (“extremely close”), then it increases the probability of next-step continuance by 3%; the weakest link (“not close”) is also the least effective because it reduces pass-along by 3% from the average. Medium-strength links (“fairly close”), however, are 1% more likely than average to pass along a chain. This finding suggests that whereas weak ties are good for increasing the span of network ties, strong ties are more effective for soliciting cooperation in a search process.

Individual attributes	Coefficients (s.e)	Probability
Age		
17 or under	0.038 (0.11)	0.01
18-29	0.14 (0.06)	0.03
30-39	0.090 (0.06)	0.02
40-49	-0.068 (0.06)	-0.01
50-59	-0.071 (0.06)	-0.02
Above 60	-0.13 (0.07)	-0.03
Education level		
Graduate school	0.18 (0.08)	0.04
College/University	0.014 (0.08)	0.0
High school	-0.14 (0.08)	-0.03
Elementary school	-0.048 (0.08)	-0.01
Income		
Very high	0.076 (0.04)	0.02
High	0.052 (0.04)	0.01
Medium	-0.0078 (0.04)	0.0
Low	-0.056 (0.04)	-0.01
Very low	-0.064 (0.05)	-0.01
Work position		
Specialist/Technical	0.028 (0.03)	0.01
Student	0.016 (0.03)	0.0
Other	0.00049 (0.02)	0.0
Unemployed/Retired	-0.0045 (0.03)	0.0
Executive/Manager	-0.040 (0.02)	-0.01
Work field		
Media/Advertising/Arts	0.098 (0.05)	0.02
Education/Science	0.059 (0.04)	0.01
IT/Telecommunication	-0.018 (0.05)	0.0
Government	-0.056 (0.05)	-0.01
Other	-0.084 (0.04)	-0.02
Fixed effects		
Intercept	-0.85 (0.12)	NA
Female	-0.063 (0.025)	-0.01
Nonwhite	-0.13 (0.041)	-0.03

Table 3.3 Coefficient estimates from a multilevel logistic regression model of next-step continuance probabilities for individual attributes. The probabilities are presented as deviation from the baseline of 30% for white males.

Relational attributes	Coefficients (s.e)	Probability
Relationship strength		
Extremely close	0.13 (0.06)	0.03
Very close	-0.013 (0.05)	0.0
Fairly close	0.05 (0.05)	0.01
Casually	-0.0093 (0.05)	0.0
Not close	-0.16 (0.07)	-0.03
Relationship origin		
Work	0.043 (0.03)	0.01
School	0.025 (0.03)	0.0
Internet	0.014 (0.03)	0.0
Mutual friend	-0.013 (0.03)	0.0
Relative	-0.028 (0.03)	-0.01
Other	-0.041 (0.03)	-0.01
Reason for choosing recipient		
Profession	0.033 (0.04)	0.01
Education	0.031 (0.04)	0.01
Work brings contact	0.020 (0.04)	0.0
Geography	-0.010 (0.03)	0.0
Other	-0.074 (0.03)	-0.01

Table 3.4 Coefficient estimates from a multilevel logistic regression model of next-step continuance probabilities for relational attributes. The probabilities are presented as deviation from the baseline of 30% for white males.

Table 3.5 displays how targets affect the probability of next-step continuance. High status targets increase the probability of passing on a chain, and participants who were assigned Target #5, who is a professor in a major research university in the Northeast, were more likely than average to continue a chain by 3%; recall that Target #5 received the most successful chains. In general, participants who had targets that were perceived as easy to reach were more likely than average to continue a chain.

Targets	Coefficients (s.e)	Probability
1	-0.063 (0.05)	-0.01
2	-0.038 (0.05)	-0.01
3	-0.053 (0.04)	-0.01
4	0.089 (0.05)	0.02
5	0.13 (0.05)	0.03
6	0.10 (0.05)	0.02
7	0.036 (0.05)	0.01
8	0.051 (0.05)	0.01
9	-0.022 (0.05)	-0.01
10	-0.014 (0.04)	0.0
11	-0.011 (0.05)	0.0
12	-0.072 (0.05)	-0.02
13	0.094 (0.05)	0.02
14	-0.037 (0.04)	-0.01
15	-0.049 (0.05)	-0.01
16	-0.14 (0.04)	-0.03
17	0.075 (0.05)	0.02
18	-0.076 (0.05)	-0.02

Table 3.5 Coefficient estimates from a multilevel logistic regression model of next-step continuance probabilities for target attributes. The probabilities are presented as deviation from the baseline of 30% for white males.

In summary, we observe that high-status participants are more likely to pass along messages to someone who will pass them along again. The absolute differences among groups are small, but because attrition rates tend to be correlated across attributes, the overall distribution of attrition rates for participants is considerably larger than is indicated by any single group effect, with attrition rates varying from 60% to 80% as shown in Figure 3.5. The distribution, however, is peaked around the mean of 70%, which means that although the attrition variation is considerably large, it is typically small. The main finding from the analysis here is that we now have evidence that attrition rates actually varied according to individual, relational, and target attributes and therefore the homogeneity attrition assumption is clearly violated. Thus, the estimator of ideal chain distribution used previously (equation 2.7) becomes

invalid; we need an estimator that can account for heterogeneous attrition. In the next section, we construct an estimator that not only accounts for heterogeneous attrition, but is also unbiased by taking into account the presence of long, but unobserved chains.

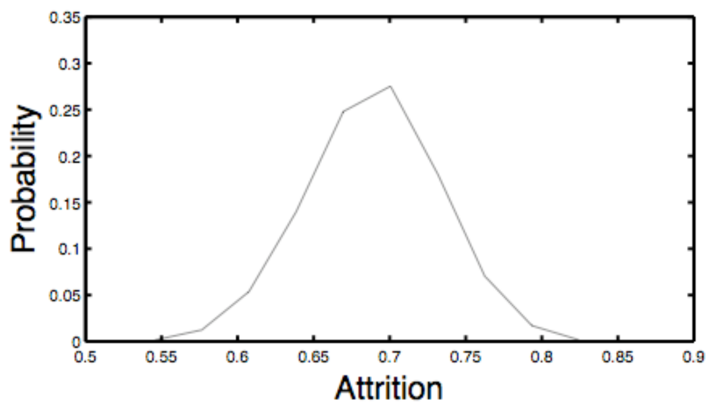

Figure 3.5 The estimated distribution of attrition over individuals. The average attrition is 0.7.

3.4 Missing data correction

The goal of this section is to construct a replacement for the estimator in equation 2.7 that can take into account heterogeneous attrition and also can be proven to be unbiased. The problem of creating an unbiased estimator is akin to the problem of missing data in statistics, so the following method that was developed by Sharad Goel (Goel, Muhamad and Watts 2009) could possibly be useful in other missing data problems. The following procedure at the most general level is similar to the estimation procedure in the previous chapter, but because it is constructed rigorously, we can be sure that the estimators are unbiased and thereby can easily incorporate heterogeneous attrition.

The basic idea is that for each completed chain that is not missing any data, we calculate the probability of observing each completed chain. Then we assign weights to those completed chains to account for the fact that some paths—the shorter paths—are more likely to have been observed than others—the longer paths. More formally, we can imagine a population of individuals and the corresponding space which is comprised of all possible paths between all pairs of individuals. In this hypothetical space, there are many possible paths connecting any two individuals, but some paths are more likely to be traced than others. An ideal small-world experiment without attrition would reveal these random paths and the lengths connecting any two individuals. Specifically, we call this hypothetical space of all possible paths Ω and a probability operator P that gives the probability of selecting a path from all possible paths between two individuals. Thus, an ideal experimental trial without attrition is akin to observing a path $\omega \in \Omega$ with probability P.

If there is attrition, and hence some data is missing, then our experimental outcomes will not always show a complete path $\omega \in \Omega$. Instead, once a path ω is drawn, it is completed with probability $Q(\omega)$ and terminated or missing with probability $1-Q(\omega)$. In small-world experiments, missing data is manifested as incomplete chains, and $Q(\omega)$ is generally smaller for longer paths, i.e., longer true paths are more likely to be observed as incomplete chains. Thus, the experimental outcome ω is observed as complete chains with probability $P(\omega)Q(\omega)$. We can sum all possible outcomes whereby a complete chain is

observed in a given trial with probability $\sum_{\omega} P(\omega)Q(\omega)$, and a missing value (incomplete chain) is observed with probability $1-\sum_{\omega} P(\omega)Q(\omega)$. Imagine we have n such trials (in our case, n is the number of chains started), and each independent trial is $X_1,\ldots,X_n$ where X_i is either a completed chain or an incomplete chain (a missing value). Technically, $X_i \in \Omega \cup \{NA\}$, where NA indicates a missing value.

Recall that our goal is to have the expected chain length without attrition, which can be stated as a weighted average over all possible paths and can be written as

$$\mu = \sum_{\omega} f(\omega)P(\omega) \tag{3.1}$$

where $f(\omega)$ is the length of the path ω. In an ideal experiment without attrition, the unbiased estimator for equation 3.1 is the usual sample average $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$. When attrition is present, however, averaging over all completed chains in the sample biases our estimates toward outcomes that are more likely to be observed, i.e., the shorter chains. This problem is severe because of the nature of the experiment in which the probability of observing a chain tends to decrease exponentially with its length. Therefore, we are much more likely to see short chains and so we underestimate mean chain length. To overcome this problem, we use the basic idea of a statistical technique called importance sampling: we re-weight samples by their inverse probability of observation to produce an unbiased estimator as stated in Theorem 1.

THEOREM 1 (S.Goel). *In the general setting described above, an unbiased estimate of the mean* $\mu = \sum_{\omega} f(\omega) P(\omega)$ *is given by*

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{m} \frac{f(X_{k_i})}{Q(X_{k_i})} \tag{3.2}$$

where $X_{k_1}, \ldots, X_{k_m}$ *are the m observed, non-missing values, and* $Q(\omega)$ *is the probability* ω *as observed uncorrupted after it has been sampled.*

PROOF. First extend f to a function $\overline{f}$ defined on $\Omega \cup \{NA\}$ (where NA indicates a missing value), and set

$$\overline{f}(\omega) = \begin{cases} f(\omega), & \omega \in \Omega \\ 0, & \omega \in NA \end{cases}.$$

Then, using the idea of importance sampling, we can rewrite the estimator as

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{\overline{f}(X_i)}{Q(X_i)}$$

where the sum is taken from all samples (including the missing values). Since the samples X_i are identically distributed, then

$$E[\widehat{\mu}] = E\left[\frac{\overline{f}(X_i)}{Q(X_i)}\right].$$

Since ω is observed non-missing with probability $P(\omega)Q(\omega)$ and $\overline{f}(NA) = 0$, we have

$$E\left[\frac{\overline{f}(X_i)}{Q(X_i)}\right] = \sum_{\omega \in \Omega} \frac{f(\omega)}{Q(\omega)} P(\omega) Q(\omega) = \sum_{\omega \in \Omega} f(\omega) P(\omega) = \mu .$$

Hence, $\widehat{\mu}$ is unbiased.

The function f in Theorem 1 is a general function; for our specific purpose we want to estimate chain length, hence we can write the unbiased estimate of chain length $\hat{\mu}$ as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{m}\frac{L\left(X_{k_i}\right)}{Q\left(X_{k_i}\right)} \tag{3.3}$$

where $X_{k_1},\ldots,X_{k_m}$ are the m observed completed chains, $L(\omega)$ is the length of chain ω, and n is the total number of chains (complete and incomplete) in the sample. We can also use Theorem 1 to estimate the entire chain length distribution that, in turn, allows us to calculate the median of the distribution. Let $f_i(\omega)=1$ if ω is a chain of length i, and $f_i(\omega)=0$ otherwise. Then, the expectation of f_i is

$$p_i = \sum_{\omega\in\Omega} f_i(\omega)P(\omega) = \sum_{\omega \text{ has length } i} P(\omega)$$

That is, p_i is the probability that a randomly-chosen chain has “true” length i. Using Theorem 1 to get an unbiased estimate of p_i, we get

$$\hat{p}_i = \frac{1}{n}\sum_{j=1}^{m}\frac{f_i\left(X_{k_j}\right)}{Q\left(X_{k_j}\right)} = \frac{1}{n}\left[\sum_{X_k \text{ has length } i}\frac{1}{Q\left(X_k\right)}\right]. \tag{3.4}$$

Therefore we can construct the estimated ideal chain length distribution by calculating $\hat{p}_i$ for each step $i=1,2,\ldots$.

The next step is to calculate the variance of estimator $\hat{\mu}$. In practice, the variability of the estimator in equation 3.3 is increased by the variability in the estimation of attrition (i.e., the error term when calculating $Q(\omega)$ by multilevel

regression). Therefore, for this source of variability we use a bootstrap sampling method, which will be discussed in the next section.

3.5. Estimating chain length distributions

In this section we are going to combine the attrition model that was constructed in section 3.3 with the new unbiased estimates developed in the previous section. We begin by analyzing data from the original Travers and Milgram experiment (Travers and Milgram 1969) followed by analyzing our data under the assumption of homogenous attrition. Finally, we use our own data to construct an estimated true chain length distribution using the heterogeneous attrition assumption.

<u>3.5.1 Homogeneous attrition</u>

The homogenous attrition model assumes that the probability of not continuing chains is the same for everyone; if this fixed termination probability is r, then the probability of continuing chains is $1-r$ regardless of the attributes of participants. Completed chains consist of a series of successful messages passing until targets are reached, and each success occurs with probability $1-r$. Thus, the probability to observe a completed chain ω with length $L(\omega)$ is $Q(\omega)=(1-r)^{L(\omega)}$. The values r and $L(\omega)$ come from the data, and then we use this expression for $Q(\omega)$ in equation 3.3 to get the estimate for true mean chain length and in equation 3.4 to get the estimate for the entire distribution of chain length.

We use the bootstrap sampling method to obtain confidence intervals for our estimates. The basic idea is to create bootstrap samples by resampling the original sample many times such that each bootstrap sample produces different estimates. Then we can construct the confidence intervals by taking the 95% range of these bootstrap estimates. Specifically, from the original sample of n chains (including complete and incomplete chains), we resample n chains with replacements to produce a bootstrap sample S_1. Some chains from the original sample are drawn to S_1 repeatedly, and some chains are never drawn at all. We repeat this resampling $k = 10{,}000$ times, producing k bootstrap samples $S_1 \ldots S_k$, where each S_i is a random resampling of the original sample. In effect, these k bootstrap samples simulate what we would have observed if we had repeated the entire experiment k times.

First, we apply the above procedure to the Travers-Milgram data in which the empirically-observed attrition rate r is 0.25. Hence, we obtain $Q(\omega) = (0.75)^{L(\omega)}$, and using equation 3.3 we get the estimated “true” mean chain length, which is 11.8 (95% CI: $8.5-15$). As a comparison, the empirically-observed mean that was calculated based on completed chains only and hence was biased is 6.2, and the longest completed chain is 11. Our unbiased estimator yields the “true” mean chain length that is both longer than the empirically-observed mean and the longest completed chain. Next, we use equation 3.4 to estimate the entire chain length distribution and we get the median of this distribution, which is 7 (95% CI: $6-7$). This median is consistent with the previous estimate by White (1970). Whereas White reported that the

median was 8, this result is actually consistent with our result here because for his calculation White assumed that senders who know the target always send the messages, and effectively set the last-step attrition probability to zero. As discussed previously, if we relax this assumption then we reduce the estimated chain length by one. Thus, our estimate of seven is actually the same as White's estimate of eight.

We use the same procedure above to estimate the "true" chain length for our data assuming homogeneous attrition. For our data, however, we differentiate the attrition for the first step in chains from the rest of the chains because we observed that the first individuals in chains have significantly lower attrition than individuals in the later stages of chains. This difference can be attributed to the fact that the participants who initiated chains were volunteers and so they were more motivated than participants who were recruited by the previous senders. There is also evidence that some messages never reached their intended recipients, and hence contribute to the higher attrition in the later steps of chains. Here a chain is completed after the first sender passed on the message with probability $1-r_0$, and then subsequent senders forwarded the messages with probability $1-r$. Thus, the probability that a chain ω with length $L(\omega)$ reaches its target is

$$Q(\omega)=(1-r_0)(1-r)^{L(\omega)-1},$$

and from the data we know that $r_0 = 0.41$ and $r = 0.70$. Plugging these numbers into equation 3.3 yields the estimate of the true mean chain length as 41.5 (95% CI: $20-68$); using equation 3.4 we get the estimated true chain length

distribution with a robust median of six (95% CI: $6-6$).

3.5.2 Heterogeneous attrition

Recall that in the attrition model, the probability of attrition is obtained by calculating the next-step continuance probability. That is, for every pair of senders and receivers, say A and B, where B is not the target, the next-step continuance probability is the probability of B to continue the chain given A's attributes (i.e., A's gender, age, income, education, work field, work position, and race), the characteristics of the relationship between A and B (i.e., the origin of relationship between A and B, the strength of A and B's relationship, and why A chose B), and the target. Thus, we can define the probability $Q(\omega)$ of observing a complete chain under the assumption of heterogeneous attrition as follows:

As illustrated in Figure 3.6, a chain ω of length $L(\omega)$ is started by ω_0 who then sends the message to ω_1 with probability $r_{\omega_0 \to \omega_1}$, then ω_1 passes it to ω_2 , and so on. For the starter ω_0 , we set the probability to pass the message $r_{\omega_0 \to \omega_1}$ to 0.59 , which is the empirically-observed value from the data. For subsequent steps in the chain, the probability to forward the message is calculated using the next-step continuance model. That is, we use ω_0 's attributes to estimate $r_{\omega_1 \to \omega_2}$. More generally, $i>1$ attributes of the $(i-1)^{th}$ participant are used to estimate $r_{\omega_i \to \omega_{i+1}}$. Then we combine each of these next-step continuance probabilities to obtain the probability $Q(\omega)$ that is given by

$$Q(\omega)=\left(1-r_{\omega_0 \to \omega_1}\right)\left(1-r_{\omega_1 \to \omega_2}\right)\ldots\left(1-r_{\omega_{L(\omega)-1} \to \omega_{L(\omega)}}\right).$$

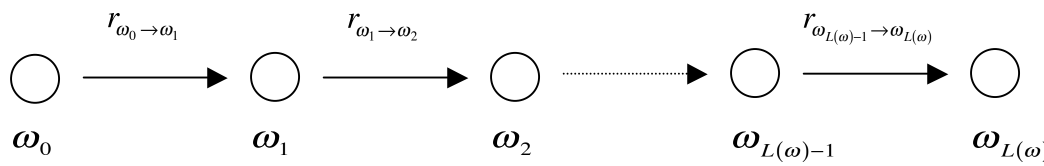

Figure 3.6 An illustration of a chain and a sequence of the probability for forwarding the message $r_{\omega_i \to \omega_{i+1}}$.

Note that the next-step continuance probability model used to estimate attrition has its own uncertainty, and this uncertainty contributes to the uncertainty in $Q(\omega)$. Therefore, this additional uncertainty must be taken into account when we construct confidence intervals for the estimates. Here we use a more robust version of the bootstrap resampling method than was used in the homogeneous model. From the original sample of $n = 162,328$ complete and incomplete chains, we resample n chains with replacements to create bootstrap samples $S_1 \ldots S_k$. These bootstrap samples can be thought of as a simulated data set had the entire experiment been repeated. Next, we need to include uncertainty from the attrition model to these bootstrap samples. Specifically, we want to incorporate uncertainty of regression coefficients from the attrition model as described in section 3.3. Usually uncertainty in regression coefficients is represented as standard error for each coefficient. In this case, however, it is useful to use the simulation method instead. The simulation generates vectors of regression coefficients $\vec{\beta}_1, \ldots, \vec{\beta}_k$ where each vector $\vec{\beta}_i$ is a complete set of coefficients for the model and comprises parameter values for each group-level effect (e.g., “college,” “18-29”). Thus, given attribute data, each vector can be used to create the estimate for next-step continuance probability. These vectors

of coefficients are generated by taking into account both the uncertainty in individual coefficients and the correlation between them (Gelman and Hill 2007). These coefficient vectors represent configurations of parameters that are consistent with the data.

For any chain ω, each coefficient vector $\vec{\beta}_i$ produces different estimates for the probability $r_{\omega_i \to \omega_{i+1}}$ and hence produces different estimates of the probability $Q(\omega)$. So we have $k = 10{,}000$ different estimates of the mean chain length while taking into account uncertainty in the attrition model; the confidence intervals for the estimates are generated by taking the range of the middle 95% of these estimates.

Using estimators 3.3 and 3.4, we find that the mean for the heterogeneous attrition model is 22 (95% CI: 4.5-57.5), and the median is 7 (95% CI: 6-8.5). Whereas the confidence interval for the mean is very wide, indicating that the estimate is very sensitive to the estimate for $Q(\omega)$, the median is more robust. The entire estimated cumulative distribution function (CDF) for chain length is shown in Figure 3.7. The variance of the CDF grows with chain length because there are few longer chains.

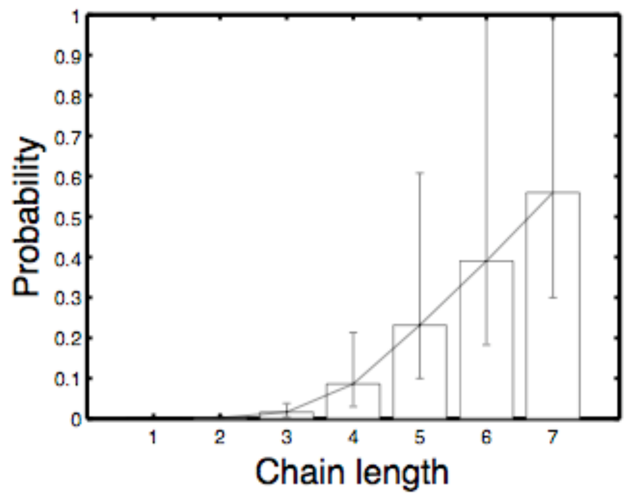

Figure 3.7 The estimated cumulative distribution of chain length under the assumption of heterogeneous attrition.

3.5.3 Randomized attrition

By definition, individuals in completed chains have passed on messages. So it is tempting to think that individuals in completed chains have lower attrition rates. In fact, individuals in complete chains are estimated to have a 3% lower attrition on average than individuals in incomplete chains (p<0.01, t-test). Consequently, one can argue that although these estimates are unbiased and have included heterogeneous attrition, the model could suffer from selection bias because individuals in complete chains are special and hence have lower attrition rates. To address this objection, we consider another heterogeneous attrition model, in which attrition probabilities R_i are randomly drawn from the distribution of estimated attrition rates as shown in Figure 3.3. In other words, each individual is independently assigned an attrition rate chosen from the population distribution. Then, we define the probability of observing a complete chain ω of length $L(\omega)$ to be $Q(\omega)=\left(1-R_{\omega_0}\right)\left(1-R_{\omega_1}\right)\ldots\left(1-R_{\omega_{L(\omega)-1}}\right)$. Using

equations 3.3 and 3.4, we find that under this randomized attrition model, the "true" mean chain length is 49 (95% CI:37-63) and the median is 6 (95% CI:6-6). Confidence intervals are generated the same way as in the heterogeneous attrition model.

To summarize, we have obtained estimates of "true" average chain length for the Travers-Milgram data under the assumption of homogeneous attrition, and for our own experimental data under three attrition assumptions: homogeneous, heterogeneous, and randomized attrition. The most striking result is that while estimates for "true" mean chain length varies widely from 11.8 to 49 and with 95% confidence interval ranging from 4.5 to 68, estimates for "true" median chain length are very robust in the range of 6-7 steps. Therefore, for about half of the population, the claim that everyone is connected to everyone else by "six degrees of separation" seems warranted. The wide variation for the mean, however, indicates that some, possibly many, chains are much longer than the median. In addition, the variation of the mean across attrition models reflects the sensitivity of chain completion to attrition rates.

Model	Mean (95% CI)	Median (95% CI)
Homogeneous attrition (Travers-Milgram)	11.8 (8.5 – 15)	7 (6 – 7)
Homogeneous attrition	41.5 (20 – 68)	6 (6 – 6)
Heterogeneous attrition	22 (4.5 – 57.5)	7 (6 – 8.5)
Randomized attrition	49 (37 – 63)	6 (6 – 6)

Table 3.6 Summary of "true" average chain length under homogeneous and heterogeneous attrition models.

CHAPTER FOUR: NETWORKING AND INDIVIDUAL HETEROGENEITIES

The main result from our small-world experiments is that the estimates of *median* chain length are robust, around six steps, but the range for the *mean* is much wider, ranging from 4.5 to 68 steps. This fact that median chain length is low but mean range is high indicates that although it is reasonable to think that social networks follow the topological small-world principle, there are individuals who cannot exploit the algorithmic small-world principle. In the real world, the inability to establish short connections to resources or information holders can have significant consequences and become a source of access inequality. In this chapter, we address the problem of how individuals can increase their efficacy when navigating social networks to get information or resources; in other words, this chapter is about networking. The term "networking" (a verb) has become even more popular than the term "network" itself. The popularity is evident as more practical business books offer advice on how to network better (for example, see Ferrazzi (2005)). A group of researchers studying entrepreneurial activities have also shown how networking by founders of new ventures greatly affect the entrepreneurial outcomes (for review, see Stuart and Sorenson (2008)).

If we look at the literature on social networks, however, networking has not been the subject of extensive research; the primary focus in this area has always been on social structure. Thus, networking is seen as an effort to achieve advantageous structural positions. For example, one networking strategy is to obtain advantaged structural positions such that information flow can easily pass

through. Research in this area has produced consistent findings that networks with diverse connections are desirable (Raider and Burt 1996; Renzulli, Aldrich and Moody 2000; Stuart and Ding 2006). Consequently, practical literature on how to “network” mostly revolves around the idea of how to build diverse non-redundant ego networks (Uzzi and Dunlap 2005). Here, the purpose of networking is to manipulate ego networks so they are well-positioned to exploit their connection to the fullest when there is a need; that is, networking as an “investment.”

Although networking as an investment is a part of what people do when networking, another type of networking can be seen as directed search. When we are searching for jobs, apartments, or investors, we construe networking as a directed search effort toward a specific resource, service, or piece of information. Seeing in this perspective, networking combines individual networking strategies, individual characteristics and the network structure in which individuals are embedded. Thus, this chapter is an effort to bridge the gap between two views of networking: networking as individual activities (directed search) and as structural positions (investment). Previous work that tried to bridge this gap is limited. Notable efforts include equilibrium analysis of networking in a labor market (Boorman 1975; Montgomery 1994), although, of course, real networking does not occur in an equilibrium condition so it is not clear how to apply results from these studies to the real world.

Here, we combined a model of macro social structure with a model of networking as individual activities. We used computer simulations to examine

various hypothetical networking scenarios, and tried to identify effective strategies and useful individual characteristics that could increase the chance of networking success. These simulations were not intended as an explanation of the algorithmic connectivity for general social networks. Rather, we took the perspective of individuals who want to increase their efficacy in networking. Consequently, there was no pressing need to run simulations for a large system, and for the individual-level analysis, even small differences in networking outcomes would matter.

There are at least two features of real-world networking that are not present in the small-world experiment. The first is the amount of information about targets available to searchers. Whereas participants in small-world experiments have to find a pre-determined target person whose complete identity is known (there was no ambiguity about who the target person is), the complete identity of the target in real-world networking is usually unknown. Instead, in a natural search process, searchers are usually looking for a type of person (e.g., a database expert, an investor in the biotech industry, or a gallery owner) rather than a specific person. In other words, in contrast to the tracing activity in small-world experiments, networking is as much about *identifying* the target as making a connection to the target.

The second difference between small-world experiments and networking is the strategy that searchers use. Participants in small-world experiments use one strategy only. Namely, each participant has to contact one of their acquaintances and use the contact who will bring the message closer to the

target. In contrast, in real-world networking, multiple networking strategies are available. For example, instead of using referrals, networking activities can be conducted through the use of an institutionalized networking event (Nohria 1992) or informal gatherings (Ingram and Morris 2007).

4.1 Networking strategies and individual heterogeneities

When networking needs to yield immediate results, however, the networking strategy would have to be aimed more directly at a target person instead of only trying to establish advantageous structural positions. Examples of direct networking strategies are referral- and non-referral-based networking. We call the referral-based networking *interpersonal* networking strategy and non-referral-based networking *targeted* networking strategy (Lee and Watts 2006). In the interpersonal strategy, one's contact introduces her to a new contact; as for the targeted strategy, one introduces herself to a stranger, presumably in a networking event.

There are benefits and costs for the interpersonal networking strategy (Vissa 2008). By definition, referrals are more knowledgeable about both searchers and targets and hence using referrals can lead to a better match, and the presence of mutual acquaintances could promote good behaviors. Referrals could also put pressure on the target to respond to the request of the searcher. There are disadvantages, however, if searches rely on referrals. As noted by Vissa (2008), referrals could operate on the logic of reciprocity, so the use of referrals could entail future obligation. In addition, because a referral, a searcher,

and a target form a triadic relationship, it puts some constraints on the nature of exchanges that can occur between searchers and targets. For example, borrowing a large sum of money from a friend or family member could put ties at risk when the debt cannot be repaid. In many searches, time pressure and the decay of trust as the length of the referral chain increases leads to very short referral chains; thus, interpersonal networking strategy limits itself to narrow opportunity space since the number and type of referrals are limited because of, say, homophily.

A targeted networking strategy can be used to open up opportunity space because searchers do not rely on referrals to make connections to anyone, even if they can connect directly to the target. The absence of referrals, however, eliminates the benefits and costs of referrals as described in the previous paragraph. When the needed information or resources are not publicly available, it would be very difficult to obtain access to them without referrals. Specifically, there is a trade-off between the specificity of the events or groups that become the target: public events or groups are easy to access, but most likely it will not be so useful for finding a target; it is difficult to find out about very specific events, not to mention getting access.

Networking strategies, however, do not determine networking success completely. As our small-world experiments have shown, individual variations also play an important role. In our experiments, the probability of completion of a search chain is sensitive to the individual attrition that, in turn, is related to individual socio-economic status and relational variables. In addition, some

individuals could also exhibit certain traits that make them better at networking. For example, searchers can develop skills to make referrals or even targets themselves be more cooperative (Hallen and Eisenhardt 2008; Zott and Huy 2007) or being persistent by initiating multiple search chains (Lee 1969).

This individual-level heterogeneity is something that has been missing in most models for small-world networks. There are models that explain how small-world structures can arise and thus how *typical* individuals are topologically and algorithmically connected (Adamic and Adar 2005; Adamic et al. 2001; Kleinberg 2000; Watts 1999; Watts, Dodds and Newman 2002), but we know little about how individuals with different characteristics can improve their effectiveness in navigating social networks. The goal of this chapter is to address this gap—how to improve individual networking success, not the average success—by studying the effect of networking strategies and individual variations on the probability of successful networking. To achieve this goal, we constructed generative models using computer simulations. Computer models are suitable for this study because we can simulate various hypothetical networking situations in different conditions and come up with testable hypotheses about what individuals can do to improve their networking activities.

In modeling networking activities, it is imperative to include the process of network formation in the model so we can gain insights on the social mechanisms that give rise to the ability to access social capital by networking. Furthermore, by explicitly including network formation process to the networking model, we can make a connection between networking activities and macro

structure (Small 2009). Thus, networking happens in existing networks that have been created independently of the networking process. This assumption also captures the notion of multiplex ties where multiple roles are embedded in one tie; the usefulness of a tie may be a by-product of another intention, e.g., women who were looking for an abortionist used ties that they never imagined to be useful for finding an abortionist (Lee 1969). In short, we want to have a model that includes the cause and the consequences of network structure.

We will explain the model for network formation in the next section, which will be followed by the model for networking in section 4.3.

4.2 The model: network formation

The model that we use for creating networks is an extension of the generalized affiliation model that was first proposed by Watts, Dodds, and Newman (2002). At its core, the model is a formalization of the Simmelian idea of duality between form and content. We can construe social structure as comprising “the objective pattern of relationships and the subjective understanding guiding relationship formation” (Martin 2009). Martin (2009) goes on to argue that the subjective understanding of relationships is related to the content of the relationship (social space, or “culture”), and the objective pattern of relationships as the form (network space, or “structure”). For our purpose here, we are not going to model network context explicitly, but it is enough to assert that social ties originated in a context within a cognitive structure, i.e., social space, and that individual perception about the structure matters.

To understand the structural origin of the small-world phenomenon, we must take into consideration the complexities implied in the following properties of social structure: identities (White 1992), cross-cutting social circles, multiple scales (Blau and Schwartz 1984), and foci of interactions (Feld 1981).

White (1992) has argued that the basic unit in social life is identity. In the context of social search, two facets of identity are especially relevant. First, identity acts as “a face” by which individuals can be identified. Here, identity can be seen as a static configuration of multiplex ties in which each individual has its own unique configuration of ties. Second, identity provides contexts for social interactions via foci of interactions (Feld 1981). The second facet is a dynamic portrayal of identity in which individuals assert their identity by participating in social interactions. Two individuals who share the same focus of interaction are more likely to form a tie than individuals who do not have a common focus of interaction, even if they have similar social attributes and positions. For example, an interest in sociology draws people from various backgrounds to study sociology and meet sociologists, which in turn leads them to become sociologists themselves and to be identified by others as such. Thus, identities induce the creation of individual attributes through interpersonal ties.

Identity operates at the group level as well. Individuals are the product of cross-cutting social circles (Blau and Schwartz 1984; Simmel 1955). The effects of cross-cutting social circles on general social structures depend greatly on the degree of the relation among social circles. If two social circles are highly correlated, then these social circles are consolidated and hinder intergroup

relations. For example, if a particular kind of occupation were completely dominated by workers from a certain race—i.e., if race and occupation were consolidated—then interracial relations based on occupation would have a low probability of occurrence. In contrast, low to medium correlation, i.e., cross-cutting social circles, increases intergroup relations. For example, in a community where there was complete segregation in the workplace but a mixed education system, then race and education would intersect with each other, forming cross-cutting social circles; thus, interracial relations are more likely to occur in an educational context. The concept of cross-cutting social circles enables us to see that the contraction of chain lengths in a small world can be caused by pairs of individuals who are very close in one social domain but far away in another.

In addition to cross-cutting social circles, Blau and Schwartz (1984) noted another important property of the social world: concentric circles. They aptly wrote,

> The components of a complex social structure are themselves social structures….Nations have provinces or states; these consist of cities and villages, which comprise neighborhoods; and each of these subunits of society has a social structure.

The nested nature of social structure implies the existence of multiple scales, and these in turn have an important consequence for the search process. Individuals may perceive that social spaces comprise many disconnected islands locally, because they have only limited knowledge about their social networks. On a global scale, however, those separated islands may in fact intersect with each other. Therefore, long-range connections are not actually long for those who are

part of them. They see them as short because of their shared memberships in the corresponding social group.

Therefore, the effectiveness of a social search is as much affected by macro social structures, such as cross-cutting social circles and multiple scales, as by individual properties such as the number of connections one has. This assertion is in stark contrast to the claim that the existence of highly-connected individuals is the necessary condition for successful searches (Barabasi 2002; Gladwell 2000). Furthermore, the hypothesis of the existence of highly-connected individuals acting as hubs still lacks empirical support (Dodds, Muhamad and Watts 2003) and is, in fact, theoretically implausible. The number of contacts of a social network hub must lie within the magnitude of the population, so the hubs can easily cover the total population in short steps; e.g., in the United States, a hub must have acquaintances in the millions. This is very unlikely since links in social networks are not maintained simultaneously (Gibson 2005).

The common feature of all four structural components described above is the focus on groups as the basis for social interactions. Consequently, our argument departs from the current literature, which focuses mainly on the property of ties in the small world, whether contractions of a chain length are caused by ties that are weak or strong, long-range or short-range, or random or not. It may be the case that most ties in overlapping regions are weak ties (Granovetter 1973), but the crucial point is the cross-cutting of social circles themselves. Weak ties can dissolve over time because of repeated interactions and transitivity (Kossinets and Watts 2005). Thus, although bridges tend to be

weak ties, short paths connecting different social groups do not require bridges. They can be strong short-range connections as long as they connect different social circles. Therefore, we argue that it is more natural to think that short paths in social networks exist as a consequence of cross-cutting social circles than to regard them as due to a random rewiring mechanism, as originally proposed by Watts and Strogatz (1998).

Within this social space, there are smaller units—i.e., categories—in which actors are embedded and can be considered as foci around which their social relations are organized (Feld 1981). Actors organize joint activities around a focus, which can include families, associations, expertise, and firms. Two actors who share the same category are not necessary connected, but they are embraced within the category in a social reference. These categories form a nested structure, i.e., each category belongs to a category of categories and so on until there is one overarching category, which is the root of all other categories.

The illustration of the model is depicted in Figure 4.1A. Actors (black dots) are characterized by a category to which they belong (gray circles). Taking a familiar example from the academic world, people are categorized according to the academic departments to which they belong. For example, in Figure 4.1A, consider *p*, *q*, *r*, and *s* as the sociology, political science, physics, and astronomy departments, respectively. Both sociology and political science belong to the larger group of social sciences, while physics and astronomy belong to the physical sciences (such larger divisions are marked by the dashed ellipse lines).

The distance between categories is defined by the lowest common ancestor (the ultrametric distance), and because categories reside in social space, this distance can be thought of as social distance. Thus, for example, the distance between the sociology (*p)* and political science (*q*) departments is one since only one step is needed to find the lowest common ancestor; the distance between the sociology (*p*) and astronomy departments (*s*) is two, which is the number of steps to the nearest branching point shared by the two categories. As can be seen from the figure 4.1A, categories form a nested structure that we call a domain, and are parameterized by the depth (*l*) and breadth (*b*) of the nested structure.

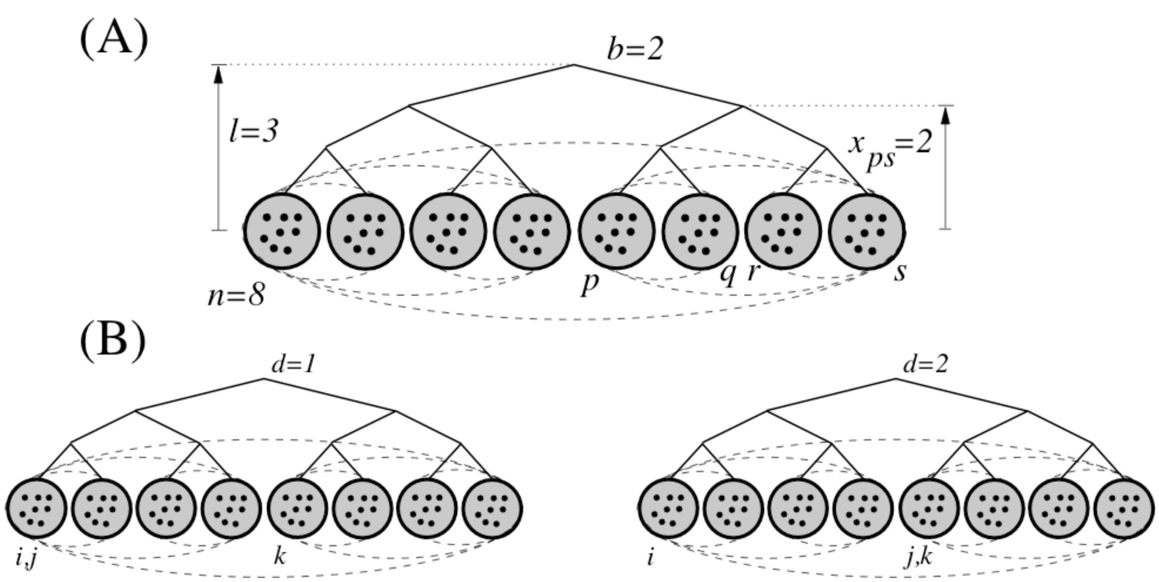

Figure 4.1 (A) Individuals (black dots) are members of categories (gray circles) and categories of categories (dashed ellipse lines) and so on, forming a tree of scales in which the whole world sits at the top of the tree. In this example there are eight individuals in each group in a tree structure with three levels (*l* =*3*) and a branching ratio *b* =2. Individuals within a group have distance zero, where distance is defined as the lowest common ancestor (i.e., ultrametric distance). For example, *p* and *s* are separated by two levels, hence their distance is two. (B) The way in which an individual can parse the world varies according to

various social domains (e.g., geography, work, and family), so the model specifies a set of trees. In this example, there are two domains, $d = 1$ and $d = 2$. The social distance across social domains is defined as the minimum ultrametric distance. Hence, this social distance can violate the triangle inequality: individuals *i* and *j* are very close to each other, but far away from *k* in domain 1; whereas in domain 2, *j* and *k* are very close but are both at a large distance from *i*.

Actors can belong to more than one domain simultaneously. Thus, we can construct other nested structures. As shown in Figure 4.1B, in addition to the domain that is based on academic discipline ($d=1$), we can have another domain, say, geography ($d=2$). Consequently, actors who are close to each other in one domain can be far away in another domain and vice versa. Taking the illustration from Figure 4.1B, we could say that two civil engineers (*i* and *j*) are very far from a sociologist (*k*) in the academic discipline domain ($d=1$), but both *j* and *k* live on the same floor of an apartment building and thus are very close to each other with respect to the geographical domain ($d=2$).

It is also possible, however, that two actors are separated by the same social distance in different domains. How actors are distributed across domains depends on the correlation among domains; in general, social domains are neither completely independent, nor completely dependent. When the correlation among domains is low, Blau and Schwartz (1984) called it “complete intersection,” that positions of actors in different domains are independent of each other. At the opposite extreme, in a condition that Blau and Schwartz (1984) called “complete consolidation,” the position of actors in one domain completely determines the positions in other domains.

We include a consolidation parameter among social domains, which corresponds to the degrees of consolidation between social circles as described by Blau and Schwartz (Lee and Watts 2006; Motter, Nishikawa and Lai 2003). To do this, first each individual is uniquely assigned to a reference domain D_{ref}. Then we construct each non-reference domain from the reference domain by swapping the positions of two individuals at distance *j* with probability $p(j) = ce^{-\beta' j}$; because shuffles happen in pairs, we preserve the size local category. The parameter β' measures the consolidation of social domains. When $\beta' > -\ln(b)$, the degree of consolidation increases and hence people who are close in one domain are likely to be close in other domains as well. For the asymptotic condition of $\beta' >> -\ln(b)$, reference and non-reference domains become identical and the model reduces to the case of one single domain. The original model in which social domains are independent is achieved when $\beta' = -\ln(b)$.

So far we have constructed a social space as a cognitive structure that can be summarized as follows. Actors belong to categories and a collection of categories that form a nested structure, and constitute a domain. The number of domains is determined by parameter *D.* Thus, the number of categories in which actors reside is determined by the number of domains, and how actors are distributed across domains is determined by the consolidation parameter. Moreover, the complete identity of an actor is the combination of all of the positions in each domain.

As we have mentioned at the beginning of this section, this social space provides subjective understanding of social interactions and becomes the basis

of network formation. In the model, we assume that the formation of network connections is driven mainly by homophily, and we measure the similarity among actors based on how similar their categories of affiliations are. Namely, the smaller the ultrametric distance between two categories, the more likely individuals in both categories are to know each other. Therefore, the probability of having a connection with distance x can be written as $p(x) \propto e^{-\alpha' x}$, where α' is the parameter that measures the degree of homophily. When $\alpha' > 0$, long-range connections will become less likely, and thus actors will tend to have relations only with those belonging to the same category; at extreme value, strong homophily could yield a world composed of disconnected cliques. At the other extreme ($\alpha' = -\ln b$) all connections have an equal probability of occurrence, generating a random network where categories are not relevant anymore.

To summarize, actors in the model are embedded within categories and a collection of categories form a nested structure that constitute a social domain. Social distance between two actors is the ultrametric distance between two categories in which actors belong. The distribution of actor positions across domains is determined by the consolidation parameter β', and actors create ties based on the homophily parameter α'.

Parameters α' and β', however, have no natural interpretation as variables and both have the range $(-\infty, \infty)$, which prevents us from using them directly for generating a set of networks within full range of homophily and consolidation. Thus, we introduce new variables α and β which describe,

respectively, the probabilities $\alpha \equiv P[x=1]$ that an individual connects to another individual in the same local category during the network formation process, and $\beta \equiv P[j=0]$ that individuals are shuffled into their reference positions on non-reference social domains. What we have done here is to map α' and β' that have range $(-\infty,\infty)$ to α' and β' that have range $[0,1]$. To do this mapping, we select values of α' and β' between 0 and 1, and then compute the corresponding values of α and β. Consequently, we have a set of values for α and β that effectively covers the entire homophily-consolidation space.

For our simulations, we constructed networks by fixing some parameters: we set the branching ratio $(b=4)$, the level $(l=4)$, the category size $(G=25)$, and the average degree $(z=24)$. Thus, each network comprised $4^{(4-1)}=64$ categories and $64\times 25=1600$ actors. After creating underlying networks, we are then ready to set up a model for networking activities.

4.3 The model: networking activities

Our model of networking activities consisted of four parts: (1) the distribution of targets, (2) a networking heuristic, (3) networking strategies, and (4) the incorporation of individual heterogeneity. Subsequent discussions will discuss each of these parts in detail.

4.3.1. Target distribution

As we have discussed, we are interested in the case where targets are a collection of actors who have the same expertise or information: for example, a doctor, a computer programmer, or a fashion editor. We operationalized this idea by assuming that targets were members of a single category within a domain, and calling this category and domain a *target category* and a *target domain* respectively. Specifically, the steps of distributing targets were as follows. We first determined the total number of targets (N_T). Then we randomly selected a target domain and followed by selecting a target category, and randomly assigned N_T actors in this target category to become targets. As we will describe in the next subsection, the selection of target category and domain affects actors' heuristic for determining the location of targets.

Consequently, although all targets belong to the same category in the target domain, they were not necessarily connected to each other. Their connections with each other depended on how connected the category was, which was determined by the homophily parameter α. Moreover, when there was more than one domain, the distribution of targets in non-target domains depended on the consolidation among domains (β). Specifically, if all domains were fully consolidated, then targets always resided within a single category in all domains. When there was some degree of cross-cuttingness among domains, however, targets who were concentrated in a single target category in the target domain were dispersed in various categories in non-target domains.

To illustrate the dispersion of targets across domains, we can imagine doing a search to find a target who is a doctor, an investor, or an actor. In the

professional domain, each of the targets is concentrated in a certain category: doctors, investors, or actors. In other domains, however, there could be differences with respect to the degree of how targets are distributed. Let's compare two domains: professions and geography. Most doctors are not distributed widely in the geographical domain, e.g., when looking for a pediatrician, it is not necessary to use the geographical cue. Investors, however, tend to concentrate in some financial centers; thus, starting a company in one of the financial hubs does make sense because it has higher number of investors in a relatively limited space. In the case of actors, geographic concentration is even more pronounced: movie actors are concentrated in Los Angeles and theatrical actors are in New York. The various degrees of concentration across social domains are captured in our model by the consolidation parameter.

The average topological distance from any actor to targets for various networks are depicted in Figure 4.2. We see that actors who were connected to at least one target were topologically close to the targets, i.e., most actors were about four steps away from targets. In summary, the qualitative behavior is that when homophily and consolidation are not too high, all actors are close to a target. At some point when both homophily and consolidation are high enough, networks start to break up, and hence the average steps to a target increases. When the homophily and consolidation are at the maximum, there are disconnected components, but for actors who are connected to a target, the distance is small. Thus, we are certain that on average, actors in the model can reach targets in short steps. However, our goal here is not to study the average

property of a network, but rather to identify networking strategies and some individual variations that can improve networking activities.

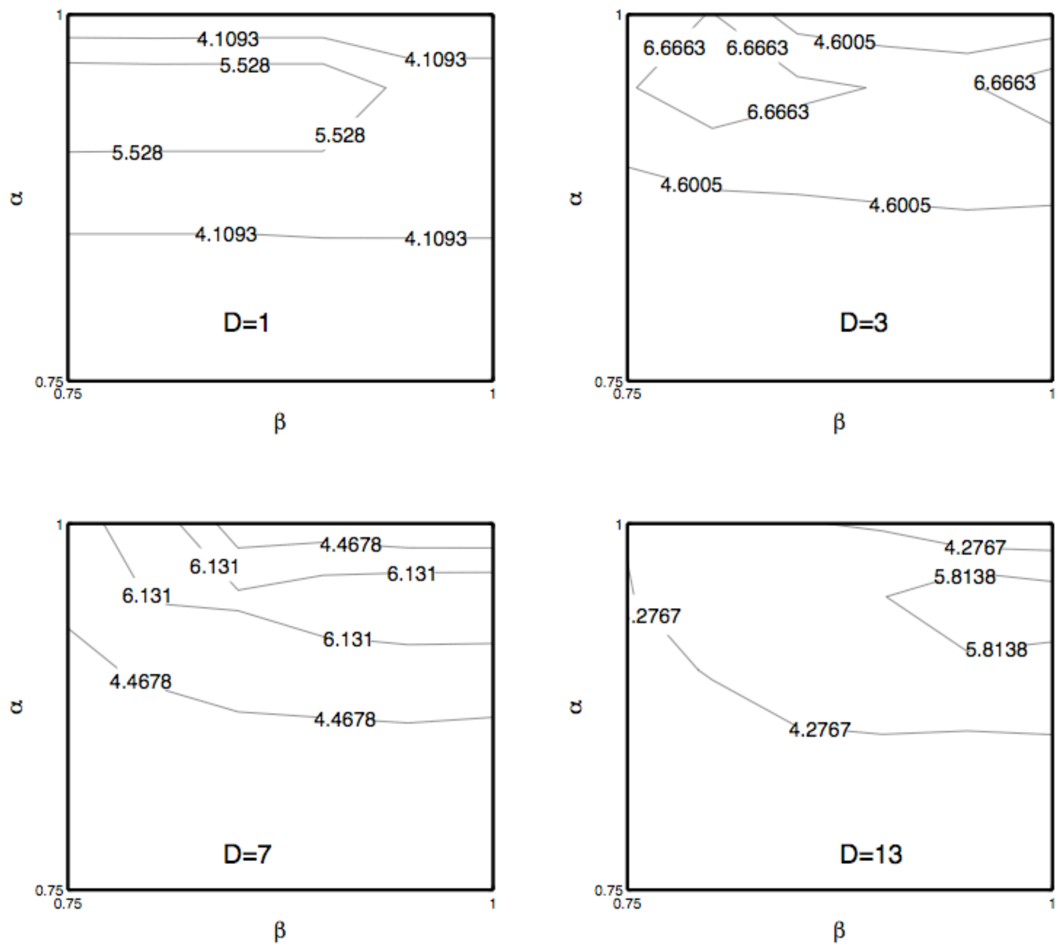

Figure 4.2 The average topological distance to targets from each actor in the network. Networks are characterized by homophily (α), consolidation among domains (β), and the number of domains (D). For each network, there are five targets, and we calculate the mean topological distance to a target for all actors connected to a target; we repeat this procedure one thousand times. We see that, on average, targets are topologically close to any actor in a network, regardless of the number of domains.

4.3.2 Networking heuristics and strategies

When networking, actors use a *networking heuristic* to estimate the location of targets, and then deploy *networking strategies* to bring them closer to targets' locations. In the context of our model, we can think of the networking heuristic as a cognitive rule for actors to identify a category within the social space in which targets were perceived to reside; we call it a *perceived target category.* Once an actor has determined a perceived target category, she then uses a networking strategy to reach it. Here we will focus on two networking strategies: the interpersonal and targeted strategies. When using the interpersonal strategy, actors use referrals to get closer to a perceived target category; when using the targeted strategy, actors form direct connections to an actor in a perceived target category.

Networking heuristics

Actors used a networking heuristic to locate targets' locations within the social space, i.e., in which category within a social domain targets were perceived to reside. We assumed that actors who are directly connected to a target could identify a target and thus make connections to a target with a certain probability. For actors who were not directly connected to a target, they used a heuristic to identify targets. We regard this heuristic to be a cognitive process operating on social space that has subjective interpretation, so it could be different from the "objective" network space. In other words, because it is possible for actors to perceive that they are far away from targets in the social

space, but actually close in the network space, networking chains do not always follow the shortest network distance to a target; this assumption tries to capture the idea from the empirical finding that people in small-world experiments do not pick the contact that would make the shortest chain to a target (Killworth et al. 2006).

We assumed that actors knew the domain of the information or resources that they were seeking, but they did not know which category within the domain was the target category. Furthermore, their ability to locate the target category depended on their social distance, measured by the ultrametric distance, to the target category: the shorter the social distance from actors to the target category within the target domain, the more precisely they could identify the target category. In other words, we can imagine that there was a cognitive constraint to locate targets in the social space; the closer actors were to the actual target category, the better their intuition in locating the target category. Thus, in this scenario, intermediaries not only passed on information to a target person, but also helped an original searcher to calibrate their search along the way.

As an illustration, consider the following situation. Imagine a consultant in a large multinational consulting firm has just been assigned to a project to assist an asset management company. At the beginning of the project, her first assignment is to find an in-house expert on, say, a derivative risk management for commodity markets. If she already knows a derivative commodity risk manager in the company, because of she has done a similar project in her previous assignment or because the expert is her friend from school, then she

can directly contact the expert. If not, then she needs to “ask around” to identify the right expert. Whom she asks depends on her knowledge. If she has no knowledge whatsoever about finance, then she will ask anyone who is perceived to be knowledgeable in finance, not necessarily a risk manager. If, on the other hand, she is already familiar with finance and risk management, then she can be more specific in her search by looking for an expert in derivative risk management, or risk management in commodity markets, or both. Intermediaries use this same heuristic, and the search process becomes more accurate as the search progresses.

Now we are ready to operationalize the heuristics in the context of our model. Actors knew the target domain, which is the same as the domain in which targets are distributed, but they did not know the target category: which was the category to which targets are distributed (see section 4.3.1). Consequently, for each networking process, actors navigated networks using only one domain: the target domain. Actors then used the networking heuristic to select a perceived target category. The accuracy of selecting a perceived target category depended on the position of the actors in the domain. If the social distance between the actual target category and an actor was one (which was the minimum distance), then the perceived target group was the same as the actual target group, i.e., the group in which targets were distributed. On the other hand, if the social distance to the target group $D_S = d,\ d > 1$, then the perceived target category was selected randomly from all categories within the distance d (see Figure 4.3). Thus, actors who were socially closer to the target had advantage because they had more

precise knowledge and were able to discriminate in a finer category, i.e., they knew “who knows what.” Actors who were in long social distance to the target, however, had difficulty in locating precisely the category in which the target belonged.

To give an illustration of how the heuristic works, let's consider the following two scenarios. In the first scenario (Figure 4.3A), an actor in the category S tries to determine the location of the target category T_c. Because the social distance from S to T_c is one, then the actor will pick T_c correctly, i.e., the perceived target category is the same as the target category. If, however, the social distance from the actor to the target category is two (Figure 4.3B), then the actor will pick—with the same probability—one of three categories (shaded categories) within that distance as the perceived target category. Thus, actors who are socially closer to targets can network with more precision.

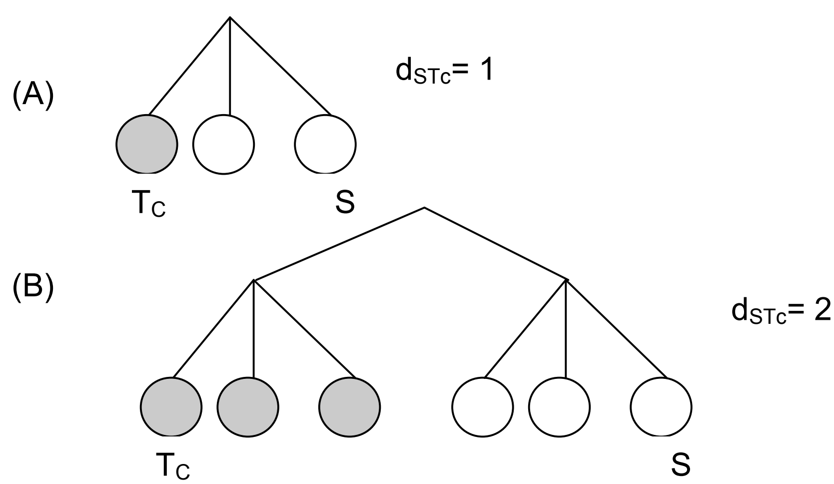

Figure 4.3 Networking heuristic. (A) Actors in the category S will pick the shaded category as their perceived target category, which is also the same as the actual target category. (B) Actors in the category S will pick one of the three shaded categories as their perceived target category. So the more socially distant an actor is from the target category, the less accurate their choice of the perceived target category.

After using the heuristic described above to determine a perceived target category, actors used a networking strategy to select a contact who is socially closer to the perceived target category.

Interpersonal networking

Interpersonal networking is the most familiar networking strategy. For example, when an employer is looking for an employee, it is rational for the employer to ask somebody who knows the employee candidate personally to get some inside information about him, e.g., whether or not he is a hard-working person, or if he has the superb skills he advertised in his paper application.

In the interpersonal networking, actors constructed a chain of intermediaries leading to a target person. At each step of the chain, an actor scanned all of his contacts, and if he found a target as one of his neighbors, then, with a certain probability, he would establish a link to the target. If, however, none of his contacts was a target person, then he chose one contact whose social distance (calculated as the ultrametric distance) to the perceived target category was smaller or the same as his.

Targeted Networking

One day in 1994, Jeff Bezos—then a Vice President in a hedge fund—was convinced that the Internet could revolutionize the book industry. His knowledge about the book industry was very limited at that time. So he attended

the American Booksellers convention to learn as much as he could about the industry. He found out that major booksellers already have electronic lists of their inventories, so all he needed to do was to put these lists in a single place on the Internet. So his next step was getting access to these lists, and he founded Amazon.com. This example illustrates how the search process itself is used to identify the target, as discussed in our networking heuristic in the previous section. In addition, the networking strategy used in this example did not use referrals, but instead used events to create new connections. We call this non-referral-based networking *targeted networking.*

Actors who used the targeted networking strategy created a direct connection to a perceived target category by selecting a random actor within the perceived target category. Thus, we can think that the networking process using the targeted strategy has no social constraints because anyone can make contact to anyone in the perceived target category. There is still, however, cognitive limitation in terms of determining the perceived target category which follows the logic of the networking heuristics described above: the closer the social distance to the target category, the more accurate actors can locate the target category.

Networking success and failure

The goal of the networking in our simulations was to establish a connection to a target. We did not limit the number of networking steps taken, but from the point of view of individuals, the actual number of steps taken mattered

because of at least two reasons: (1) for the interpersonal networking strategy, we can imagine that the efficacy of referrals decreases as the number of intermediaries in a chain increases; (2) chains that are too long would take too much time to complete, and this is especially problematic in a competitive environment. Thus, although we let chains grow unrestricted, we will focus on chains whose length is not too long because for an actor those chains are still feasible to traverse.

In our simulations, there were three reasons for chains to be terminated: individual attrition, no more unvisited actors, and chains could not move closer to the target category within the social space. We assigned each actor an attrition probability to continue a chain, and hence chains could be terminated because of attrition. We also did not allow actors to participate in networking activities more than once; thus, loops were not possible.

In our experiment, we observed that Target #5 had the lowest average attrition rate. We think that his position as a university professor in a large research university in the Northeast created the perception that the target was easy to reach, so the attrition was lower. We incorporate this idea that mental maps about the target affects search process by assuming that actors would not continue a chain if they could not find a contact who was socially closer to the target category. This assumption made explicit the linkage between actors' subjective perception about social space and networking outcomes.

4.3.3 Individual heterogeneity

Our main interest here is to examine the effect of individual variations to networking success. We wanted to isolate the individual-level characteristic that renders higher networking success probability. We chose to examine the following individual variations: degree, attrition, proximity, skill, and persistence.

Degree represents the number of contacts actors have. We created networks with a fixed degree distribution that is uniform, and all have the average degree of 24. To analyze the effect of the degree one has, we compared networking performance among those within the bottom and top 10% of the degree distribution in each network.

Attrition is the probability that an actor continues a networking chain. We took the empirically-observed attrition distribution (Figure 3.3), a normal distribution with $\mu = 0.7$ and $\sigma = 0.04$, and assigned randomly-drawn attrition values from this distribution to each actor in the model. Then we compared the performance of those in the bottom and top 10% from the attrition distribution.

Proximity measures the topological distance to the closest target. Our intention was to make a direct comparison between topological and algorithmic distances.

Skill is a measure for the effectiveness of persuading someone to help by continuing a search chain. We modeled skill by reduction of attrition. That is, we

divided the population into two categories, skillful and unskillful actors; skillful actors could reduce their own *and* other actors' attritions to zero.

Persistence reflects actors' efforts to start a new chain whenever a chain fails to reach its target. In our model, actors with persistence would restart the networking process until there were no more actors to contact.

4.4 Results

The following results are from simulations in which we used different values of homophily (α) and consolidation (β), fixed the number of domains D to three and repeated each simulation one thousand times.

4.4.1 Attrition

Recall that each actor in the model was randomly assigned an attrition value drawn from the empirical attrition distribution (Figure 3.3). We then compared the percentage of networking success rates between starters who had low attrition rates (bottom 10% of the attrition distribution that was an attrition rate of 65% or lower) and actors with high attrition rates (top 10% of the attrition distribution that was an attrition rate of 75% or higher). Thus, we compared attritions for first links only and focused our analysis on extreme cases. As we will see, doing extreme cases is enough because even differences in attrition for extreme cases do not matter. In Figure 4.4, these two groups of starters are displayed on the x-axis as "Low" and "High" respectively.

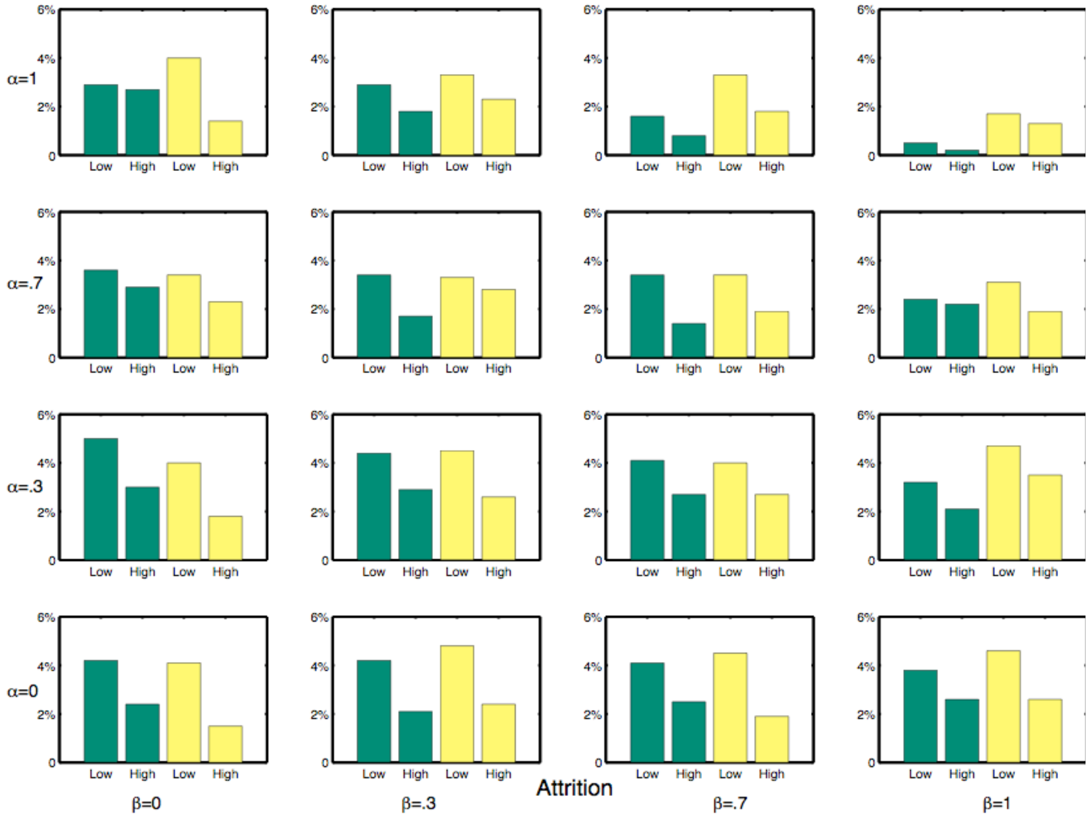

Figure 4.4 Percentages of search success for starters with low and high attritions using the interpersonal (dark gray) and targeted (light gray) strategies. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

At first glance, we see that starters with low attritions are slightly more successful in finding targets than high-attrition actors. If we look at these differences closer, however, low-attrition actors are only 2% more successful than high-attrition actors using both networking strategies. Although actors using the targeted networking strategy are not constrained by network structure, they are still greatly affected by attrition as they need the cooperation of intermediaries to identify and locate the right target category. Therefore, for both

networking strategies, these results suggest that what matters most is not so much the attrition of actors who start chains as much as cooperation from subsequent intermediaries in the chains.

The second result that we can infer from Figure 4.4 is about the effects of homophily and consolidation to networking success. First, the proportion of networking success seems to be stable in the $\alpha - \beta$ space, except in the extreme condition of complete homophily and consolidation. The effect of high homophily to networking ability is not so pronounced when the degree of consolidation among domains is low. As domains become more consolidated, however, the proportion of networking success is reduced, especially when homophily increases to the maximum.

It is expected that high α and β will reduce the efficacy of the interpersonal networking strategy. High homophily means that actors are much more likely to form ties within the same category, and this tendency in one domain is replicated in other domains as domains are highly consolidated. Thus, it is harder for actors to break away from their original category using their social ties. In the case of targeted networking, however, actors forgo social connections and make random connections with someone in a perceived target category, so high homophily should not hinder the targeted networking process. Furthermore, because categories tend to be completely connected in the situation of complete homophily, the chance of finding a target in the target category is very high. Yet, as Figure 4.4 shows, when $\alpha = 1$ and $\beta = 1$, the percentage of success for the

targeted networking strategy is also reduced, although not as much as in the case of the interpersonal strategy. So what's going on?

The problem here is that actors do not always know exactly the position of the target category. The condition of complete homophily means that one's friends are concentrated in the same category, so all targets' friends tend to be concentrated within the target category. Because domains are completely consolidated, this concentration of targets and their first-degree friends within the target category occurs in all domains. Thus, actors have to make direct connections to the target category to find a target. When homophily is lower, we can expect some of targets' friends to be in other categories than the target category, so it is still possible to find a target without making direct connections to the target category. In the case of complete homophily and consolidation, however, the only way to find a target is to create connections to the target category. Because actors who are socially distant from the target category still need intermediaries to locate the right target category, attrition rates from these intermediaries affect the networking process. Hence, the efficacy of the targeted networking is also reduced when $\alpha = \beta = 1$.

Thus, we see here that attrition makes the presence of long chains less likely, so actors can only establish connections to targets who are already topologically close to them. The average length of a completed chain here is 1.6. It is encouraging to see that the overall qualitative result from our model is consistent with the empirical finding that most chains do not reach their intended targets, but those that are completed do so in short steps.

4.4.2 Degree

Next, we analyze the effect of the number of contacts (degree) on the percentage of networking success. Actors in our models have 24 contacts on average, and we divided populations into low- and high-degree actors. Low-degree actors are those whose degree is on or below the bottom 10% of the degree distribution, and high-degree actors are those whose degree is above 90% of the degree distribution.

Results are shown in Figure 4.5 and we see there is no clear pattern emerging: across homophily and consolidation space, neither low-degree nor high-degree starters are consistently better or worse. Because the percentage of success is around the average of about 3% except for very high homophily and consolidation, we suspect that the variation in the proportion of success arises from the variation of attrition and network parameters α and β. Thus, our results suggest that the number of friends has little affect on one's effectiveness in networking.

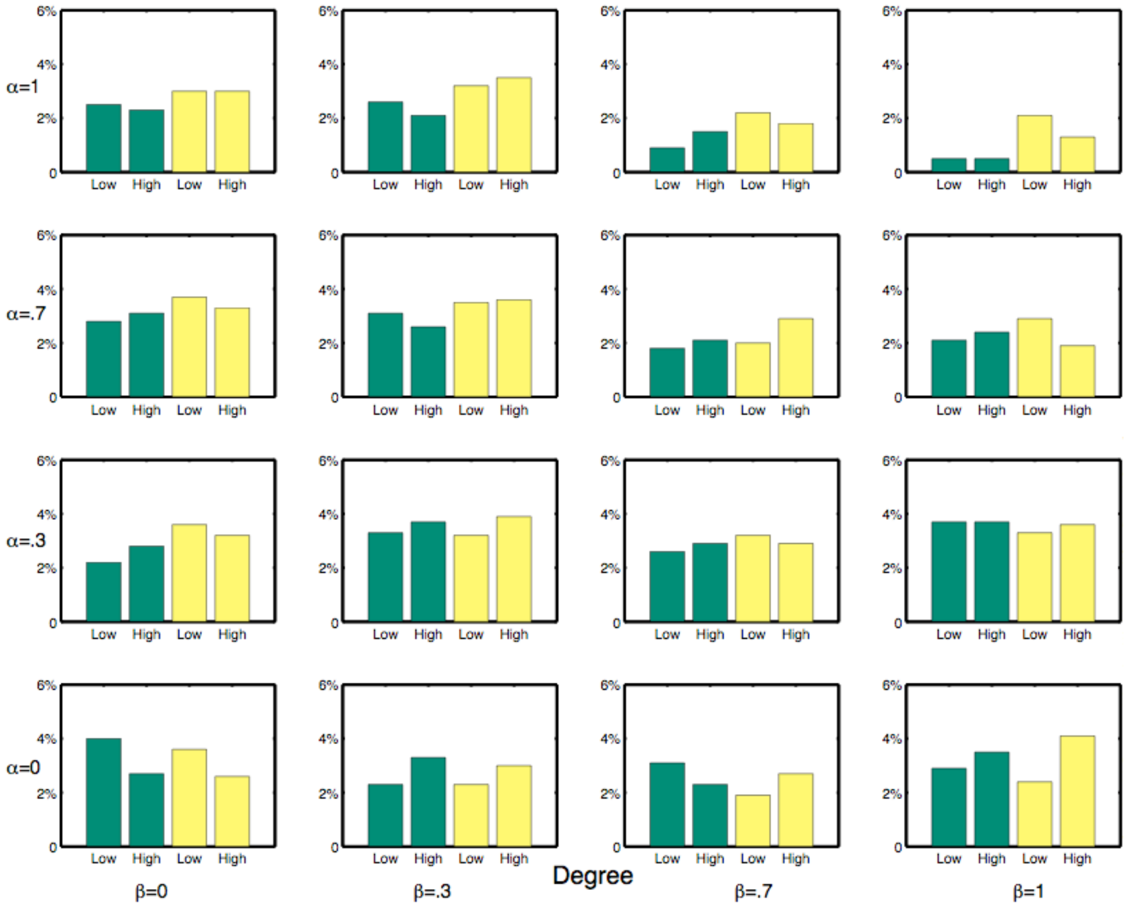

Figure 4.5 Percentages of networking success for starters with low and high degrees using the interpersonal (dark gray) and targeted (light gray) strategies. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

4.4.3 Persistence

Whereas actors without persistence start only one networking chain, persistent actors start multiple chains. In the case of the interpersonal strategy, actors started multiple chains until either a target was found or all neighbors were visited. In the case of the targeted strategy, the networking process ended only

when a target was found or there were no more nodes to contact. Results are shown in Figure 4.6.

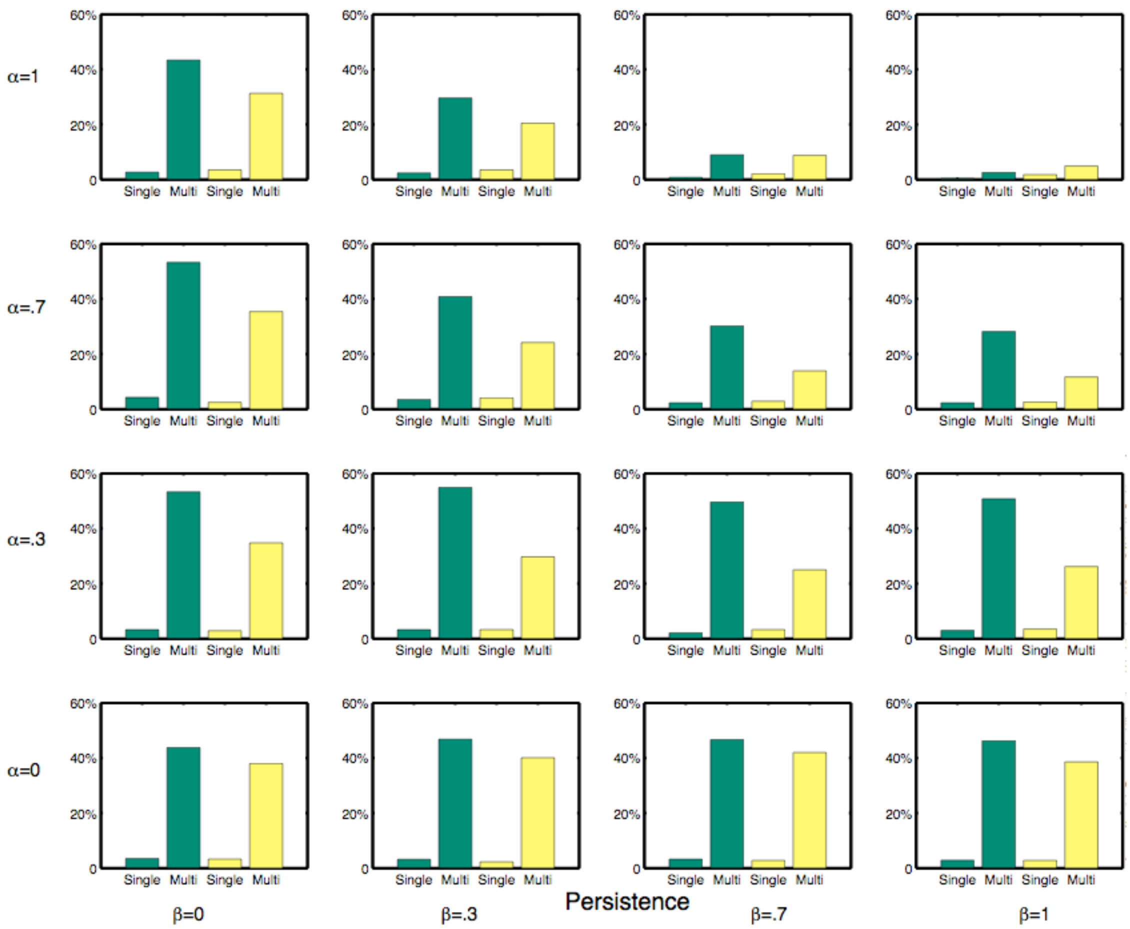

Figure 4.6 Percentages of search success for actors with single (unpersistent) and multiple (persistent) starts using the interpersonal (dark gray) and targeted (light gray) strategies. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

The general pattern here is that persistent actors are overwhelmingly more successful than non-persistent actors, and the interpersonal strategy is superior to the targeted strategy for most regions in the homophily-consolidation space. Persistent actors using the interpersonal strategy are more successful

because not because actors can reach more distant targets, but because actors exploit all connections to find the closest target. For example, when homophily is zero, we basically have random networks; hence targets are topologically close to anyone, so persistence really pays off. However, as homophily and consolidation become higher, the effectiveness of persistence decreases. This is because networks become more clustered, so actors need to break away from their cliques to find a target and chains tend to be longer. Because of attrition, longer chains are still less likely to be completed even for persistent actors.

In the case of the targeted strategy, actors focus on social space and their subjective interpretation of it. Because there are some correlations between social and network space, persistently establishing new connections based on social categories increases the chance of success, although not as much as when actors simply enumerate and use their existing contacts one by one.

4.4.4 Skill

Now we compare networking outcomes among actors who had strong networking skills and those who had no skill. Networking skill was implemented as the ability to reduce both their own and other actors' attritions to zero. Thus, for skillful actors, networking outcomes are determined by networking strategies and network structure. Figure 4.7 shows the result. Not surprisingly, skillful actors' networking performances are much better than those without skills. Yet, the success of actors with skill also depended on their strategies and the amount of homophily and consolidation.

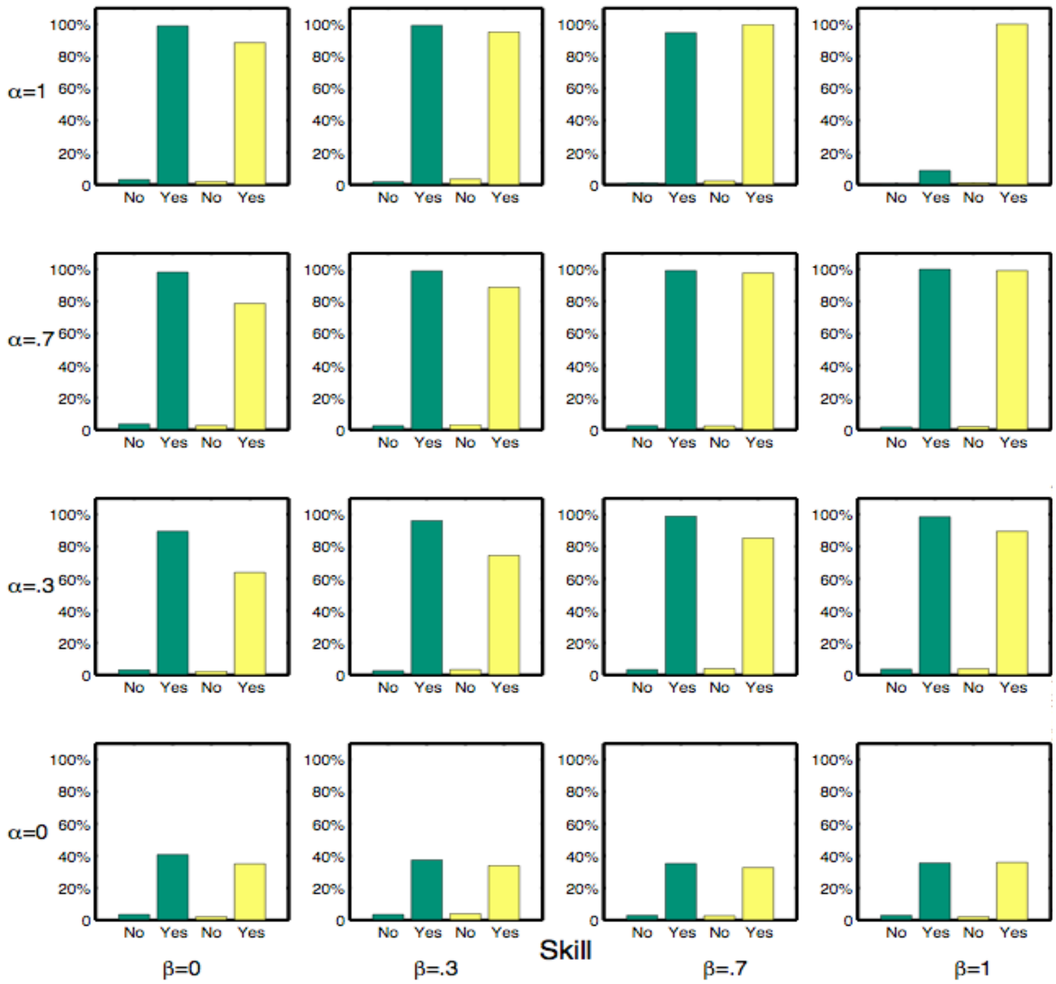

Figure 4.7 Percentages of search success for actors without and with skill using the interpersonal (dark gray) and targeted (light gray) strategies. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

When homophily is absent, both interpersonal and targeted strategies yielded the same results regardless of domain consolidation; the percentages of networking success, however, were less than 40%, which was lower than when homophily is present. When $\alpha = 0$, the dominant cause for termination was that there are no available categories that are closer to the target category. In some

cases, target category is reached but no target is found. One plausible explanation is that the absence of homophily rendered the networking heuristic useless because targets are scattered randomly in the social space, and so the location of the target category is irrelevant from the position of targets. As homophily increases, cues from the social space become more useful and hence increase networking success. If homophily and consolidation are at the maximum, however, clustering become more pronounced to the point that networks exhibit disconnected cliques, so the interpersonal strategy becomes less effective; on the contrary, the targeted strategy performs well in this extreme condition.

Next, we varied the level of skill and examined its effects on the proportion of networking success and the mean of chain length for both networking strategies. We assigned a parameter for skill $s = 0,\ldots,1$ for each actor. For example, when $s = 0$, actors' attritions stayed the same, when $s = 0.5$, actors' attritions were reduced to 50% of their original values, and when $s = 1$, actors had zero attrition.

The effect of increasing skill to the proportion of success for the interpersonal strategy is depicted in Figure 4.8. The shape of the curves are not linear where very skillful actors have disproportionate advantage to less skillful actors, and increases in skill give increasing return in terms of success rate. When homophily is very low, even actors who are extremely skillful ($s = 1$) cannot achieve 100% success rates because the lack of homophily renders finding the target category very difficult. The success rate also drops when homophily and

consolidation are too high, resulting in disconnected networks. From Figure 4.9, we see that the mean of chain length increases as actors became more skillful. This suggests that skill is beneficial because it allows actors to construct long chains and hence increase their reach.

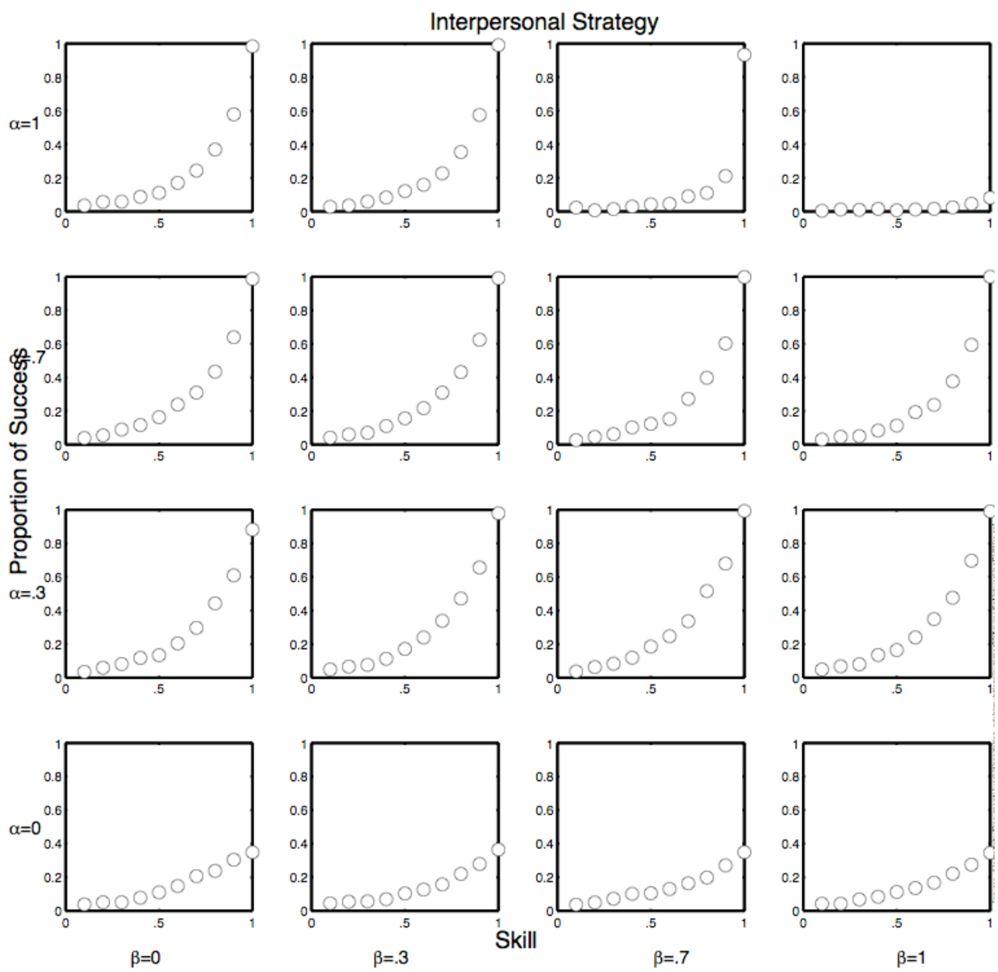

Figure 4.8 Interpersonal networking. Proportion of networking success as a function of actors' skills. Skill is modeled as attrition reduction. For example, if an actor's skill is 0.5, then their own attrition and the attrition of any other actor with whom they interact are reduced to half of the original attritions.

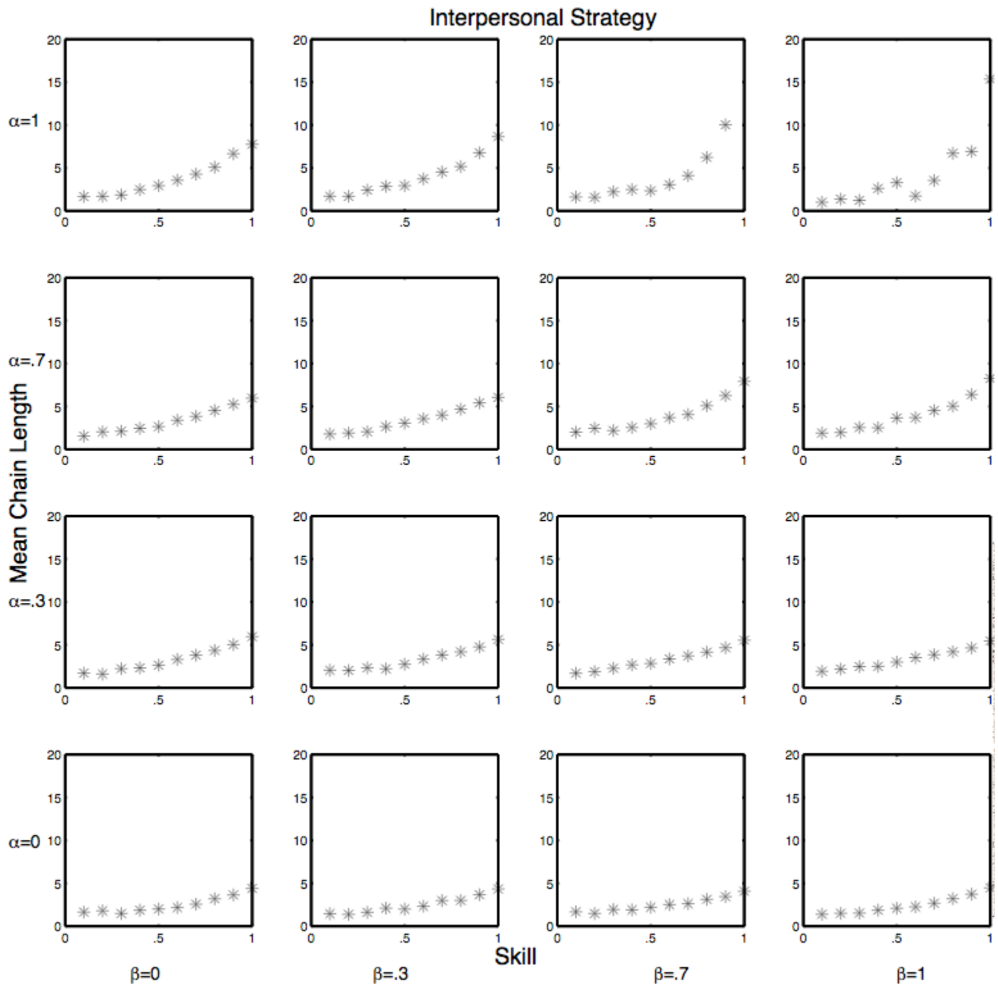

Figure 4.9 Interpersonal networking. Mean chain length as a function of actors' skills. Skill is modeled as attrition reduction. For example, if an actor's skill is 0.5, then their own attrition and the attrition of any other actor with whom they interact are reduced to half of the original attritions.

In the case of the targeted strategy, Figures 4.10 and 4.11 show how skill affects the proportion of networking success and mean chain length, respectively. For the targeted strategy, success rates show no difference in the interpersonal strategy in the region of low homophily. Unlike for the interpersonal strategy, however, the effect of consolidation is more pronounced as it increases success rates, especially for those with very high skills. As for the average number of steps, the targeted and interpersonal strategies yield similar results,

except in the extreme case of high homophily and consolidation where the targeted strategy could accomplish the task within relatively short steps. We also note that the mean chain length when there is no homophily tends to be shorter than when homophily was present. Again, the reason could be because the networks are simply random networks in which targets tended to be close.

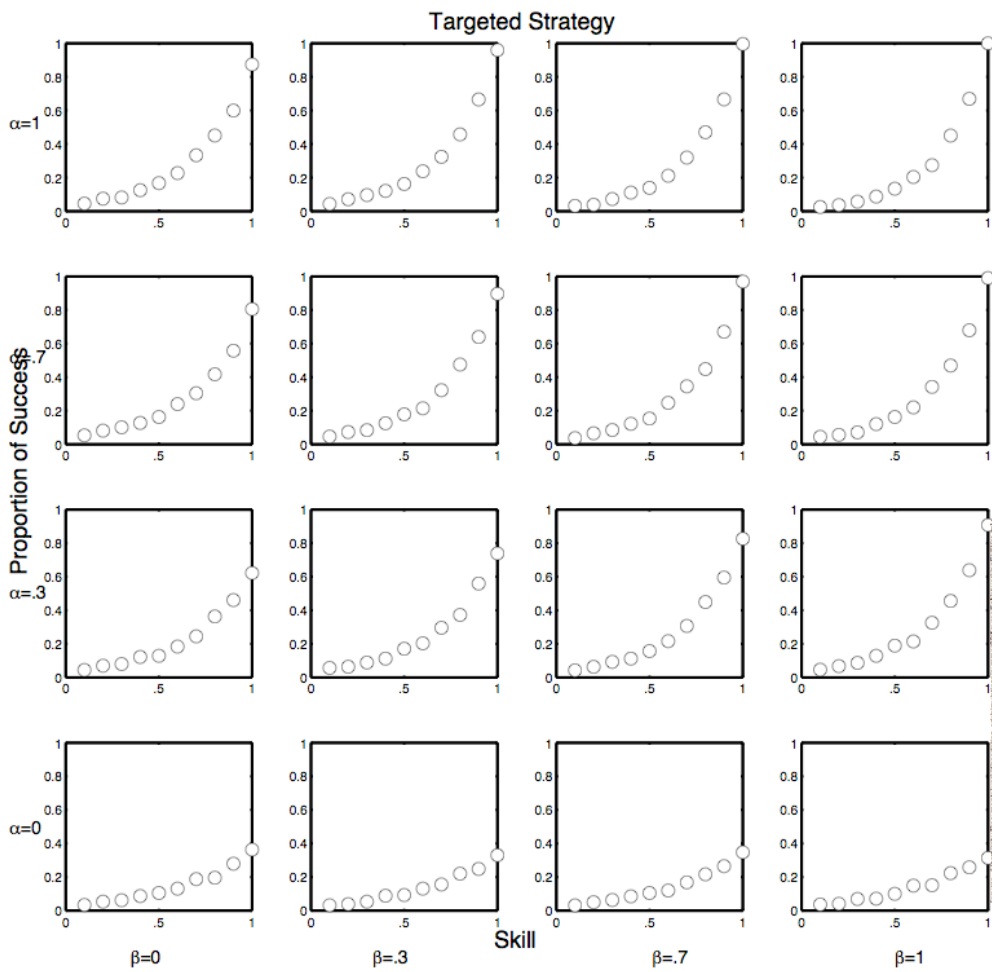

Figure 4.10 Targeted networking. Proportion of networking success as a function of actors' skills. Skill is modeled as attrition reduction. For example, if an actor's skill is 0.5, then their own attrition and the attrition of any other actor with whom they interact are reduced to half of the original attritions.

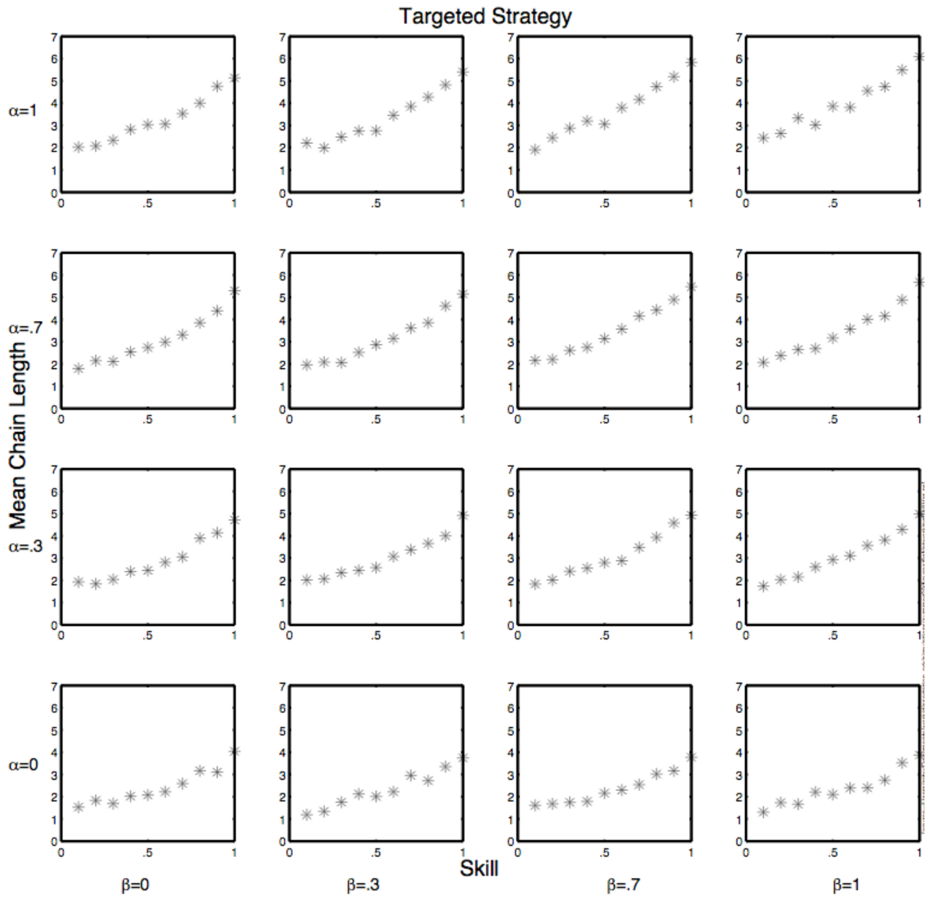

Figure 4.11 Targeted networking. Mean chain length as a function of actors' skills. Skill is modeled as attrition reduction. For example, if an actor's skill is 0.5, then their own attrition and the attrition of any other actor with whom they interact are reduced to half of the original attritions.

4.4.5 Topological proximity

For the last analysis, we turn our attention to the relationship between topological and algorithmic distances to targets. For the following simulations, we categorized actors based on their topological distance to the nearest target and performed one thousand trials for each distance. For example, we took all nodes within one step and performed one thousand networking simulations, followed by all nodes within two steps away, and so on. Thus, for each network, the total

number of trials was one thousand multiplied by the maximum topological distance in each network, which ranged from three to nine steps.

When there is a target that is one step away, success is determined solely by attrition. However, if the closest targets are two or more steps away, in addition to attrition, actors' networking strategy, heuristic, and macro structure start to matter. First, we look at success rates for the interpersonal strategy (Figure 4.12) and the targeted strategy (Figure 4.13). We observe that most completed chains are short chains. This result is not surprising because attrition accumulates and renders long chains, which have a very low probability of completion. From both figures, we also see that the proportions of success drop significantly from the first to second steps. There is one exception, however, in the case of maximum homophily and consolidation where the targeted strategy produces more networking successes. In this extreme case, there are disconnected cliques, but targets are concentrated in a category. Hence, the targeted strategy performs better than the interpersonal strategy because the targeted strategy does not require social ties to connect to a target.

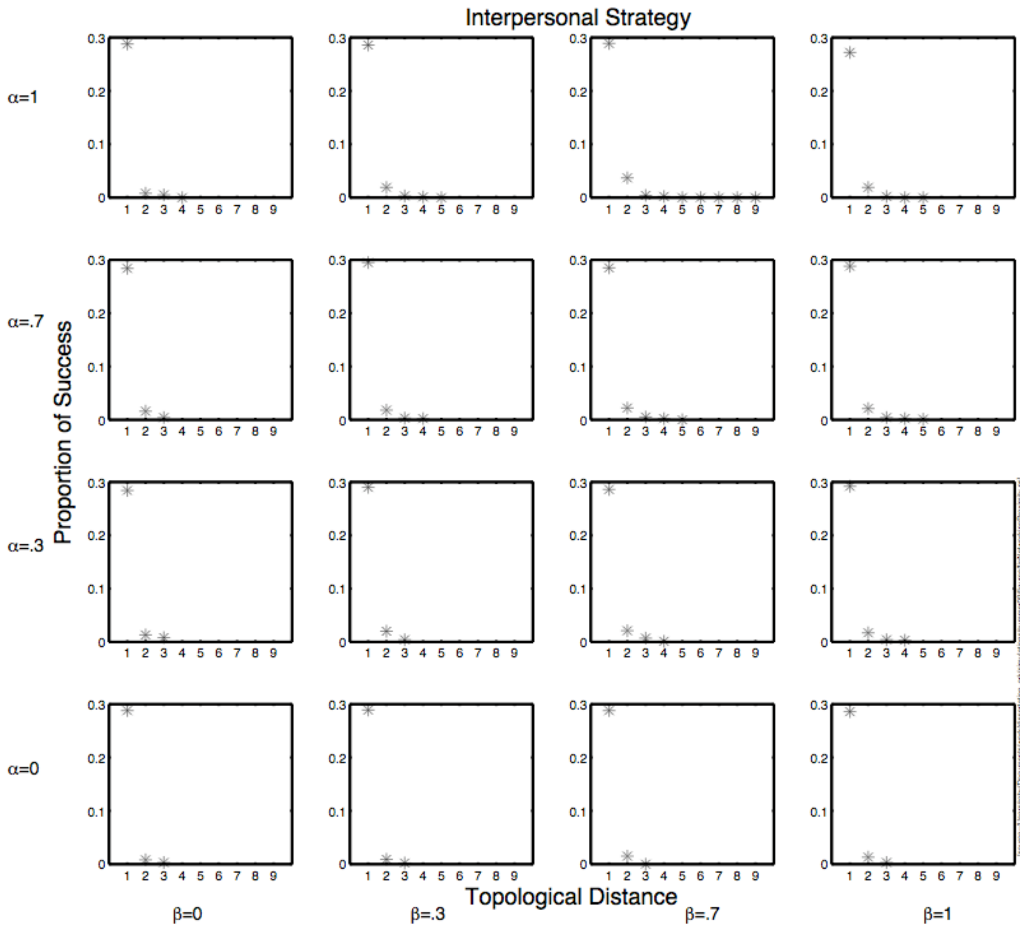

Figure 4.12 The proportion of networking successes for the interpersonal strategy as a function of topological distance to targets. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

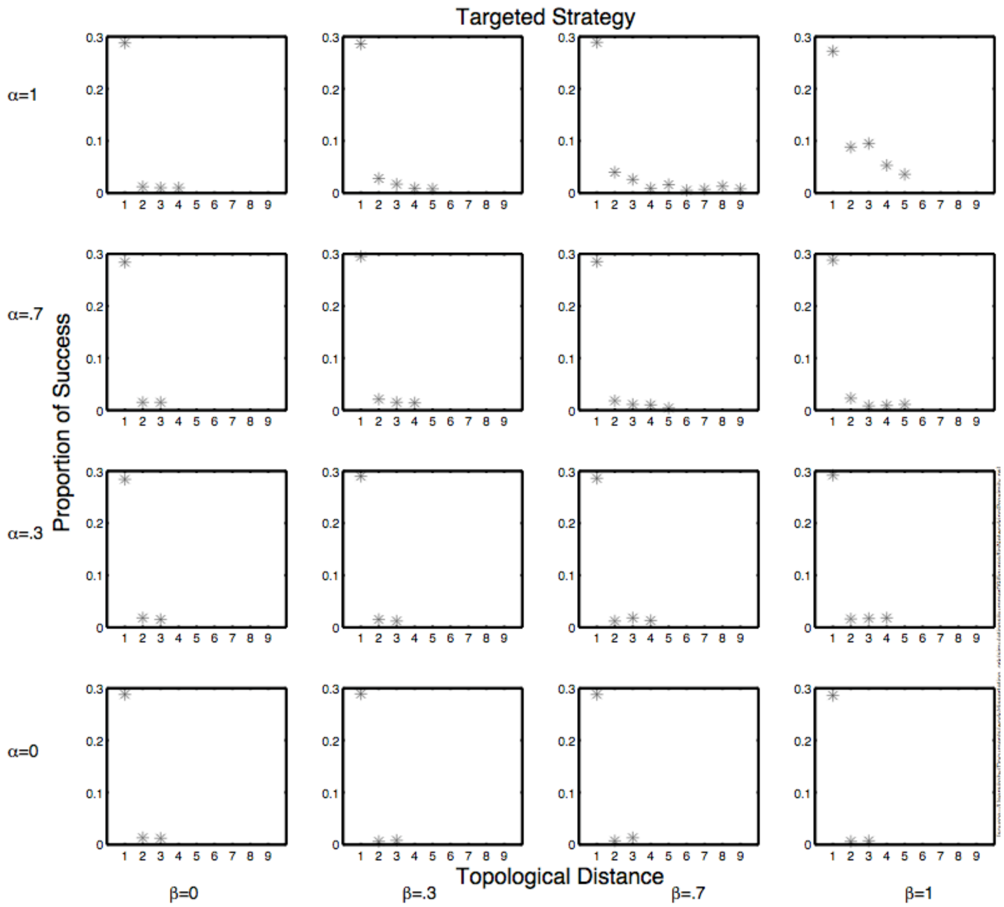

Figure 4.13 The proportion of networking successes for the targeted strategy as a function of topological distance to targets. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

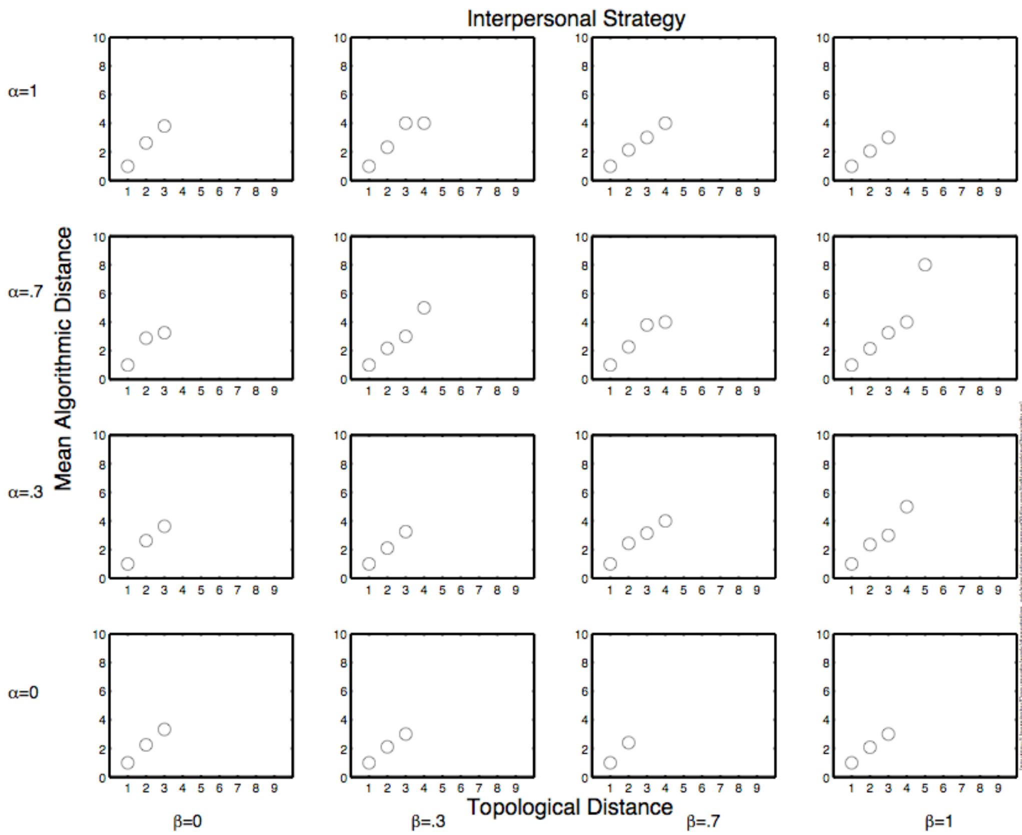

Figure 4.14 Algorithmic distance as a function of topological distance for the interpersonal networking strategy. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

Next, we plot algorithmic distance versus topological distance for the interpersonal (Figure 4.14) and the targeted (Figure 4.15) strategies. When looking at these two figures, however, we have to keep in mind the previous results about the proportion of success that shows there are only few data points for chains longer than two steps. Thus, although we observe linear relationships between algorithmic and topoogical distances, we cannot infer with confidence that these relationships are valid in general.

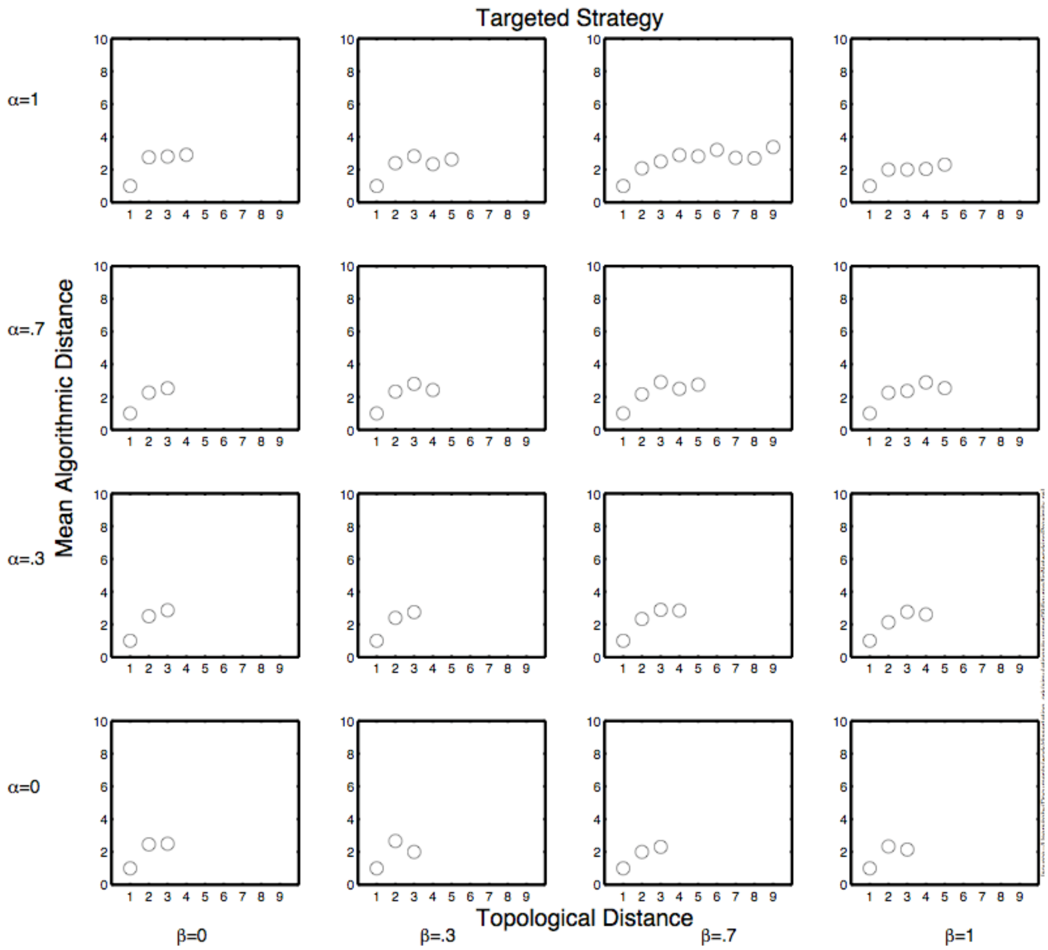

Figure 4.15 Algorithmic distance as a function of topological distance for the targeted networking strategy. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

To investigate further how topological distance affects networking, we ran networking simulations in the case of zero attrition. For the interpersonal strategy (Figure 4.16), again we observe that the absence of homophily rendered the interpersonal strategy ineffective. To be effective, the interpersonal strategy needs relatively high homophily (about 0.7, see Figure 4.16) so the success rate reaches above 90% regardless of the level of consolidation among domains. If the level of homophily is maximum, however, as consolidation increases,

success rates fall because of disconnected components. For the targeted strategy (Figure 4.17), although the proportion of networking successes is high, it is lower than the interpersonal strategy, especially for actors who were more than one step away from a target.

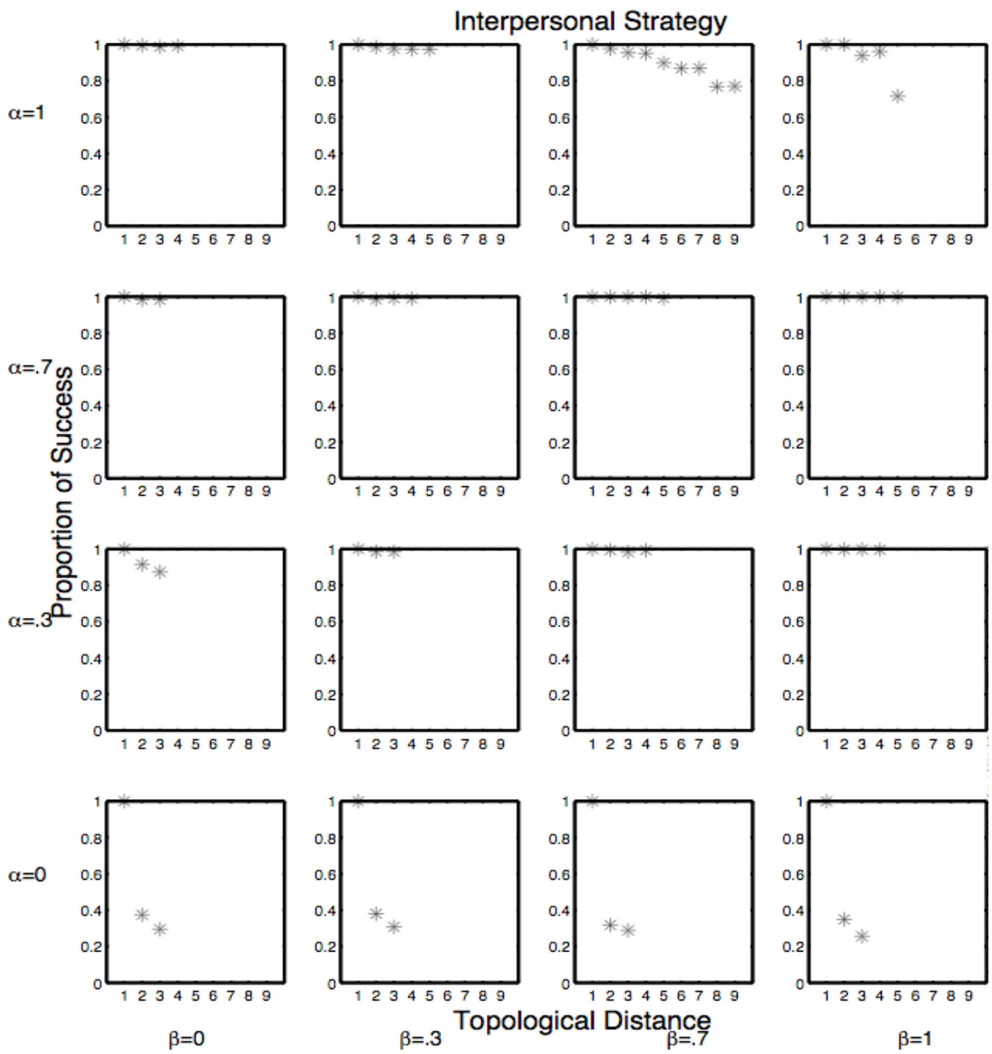

Figure 4.16 The proportion of networking successes for the interpersonal strategy as a function of topological distance to targets without attrition. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

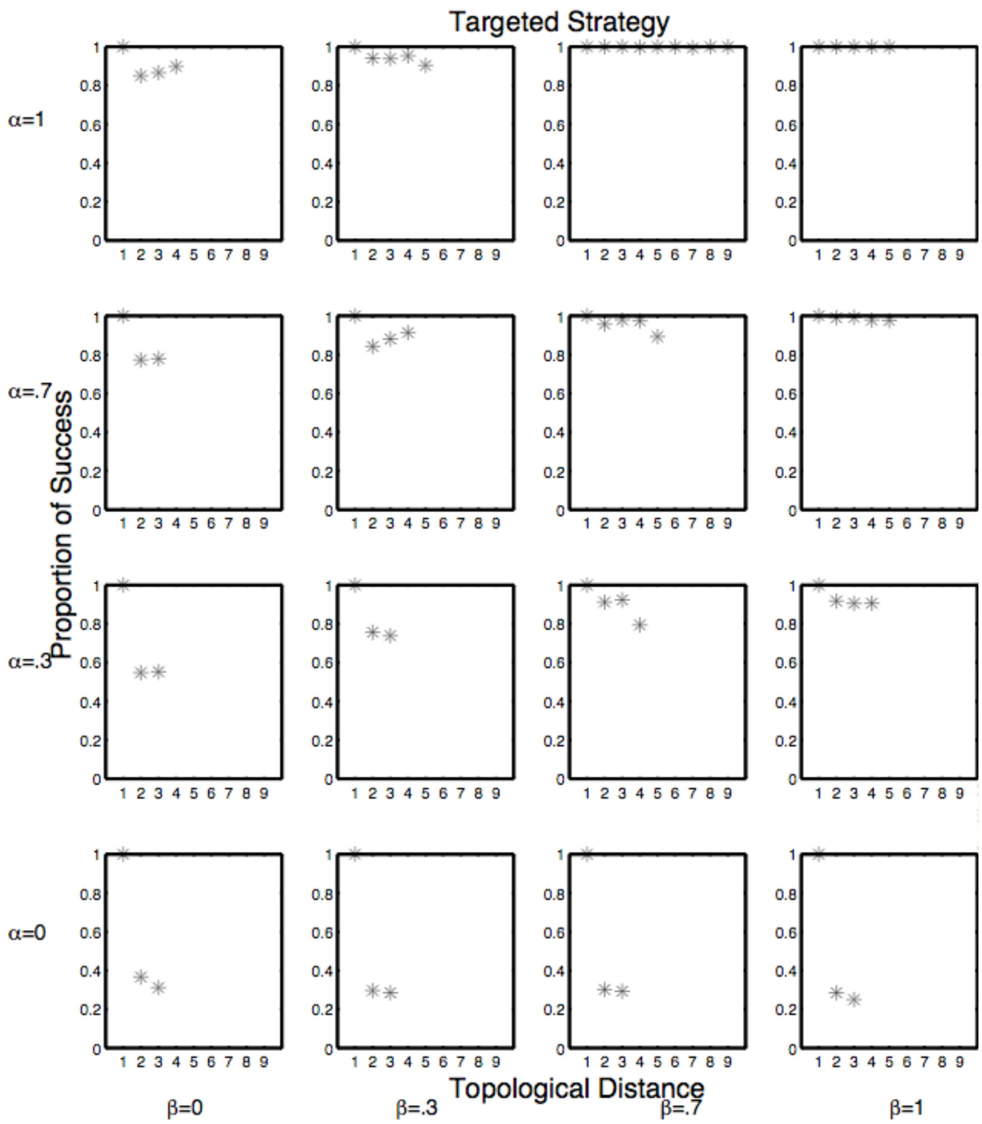

Figure 4.17 The proportion of networking successes for the targeted strategy as a function of topological distance to targets without attrition. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

Although actors have high success rates reaching targets when attrition does not exist, actors who are topologically distant from targets are still at a disadvantage. For example, using either the interpersonal strategy (Figure 4.18) or the targeted strategy (Figure 4.19), actors who are two steps away from

targets require about five steps to reach a target. For the interpersonal strategy, as topological distance increased, the mean algorithmic distance increased nonlinearly. In contrast, for the targeted networking strategy the mean algorithmic distance seems to reach a plateau after a couple of steps. Note that when both homophily and consolidation are very high, the algorithmic distance for the interpersonal strategy grows very fast as the topological distance increases (upper right corner in Figure 4.18).

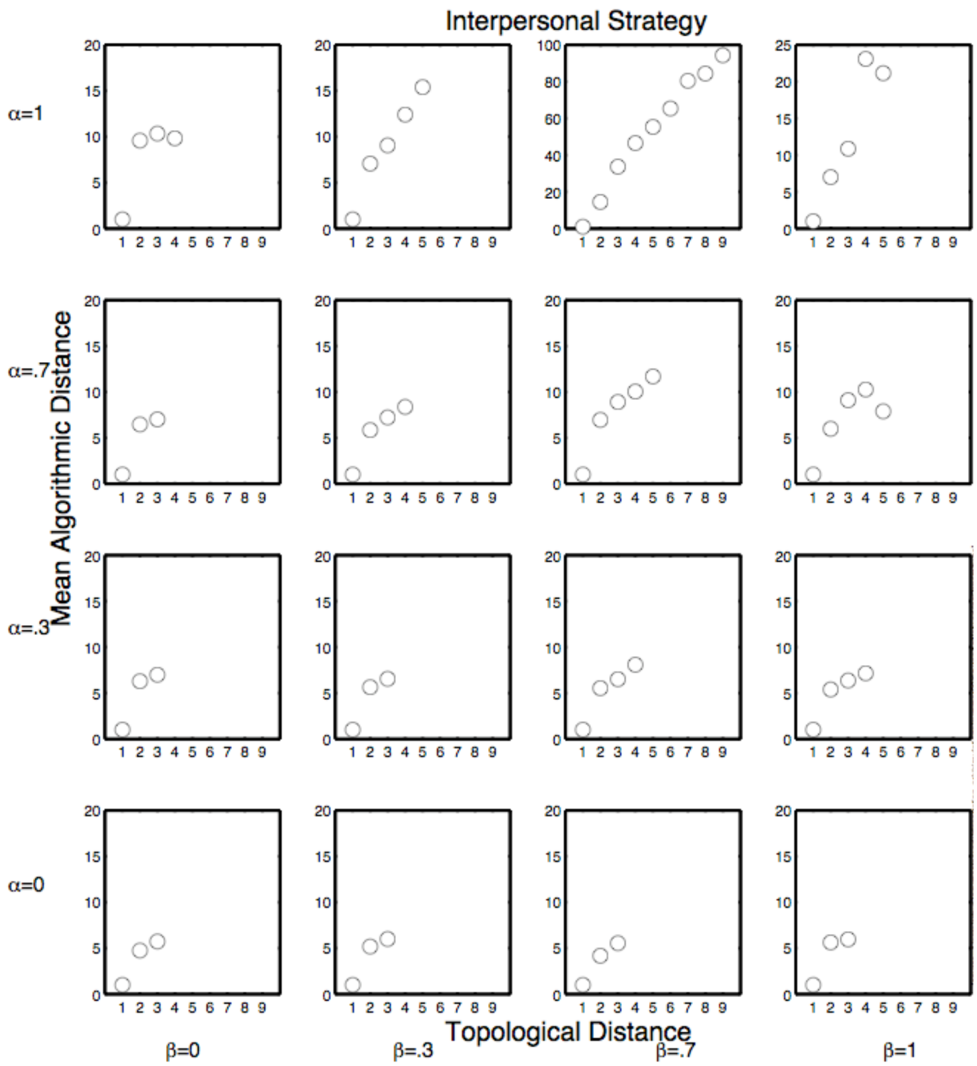

Figure 4.18 Algorithmic distance as a function of topological distance for the interpersonal networking strategy without attrition. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

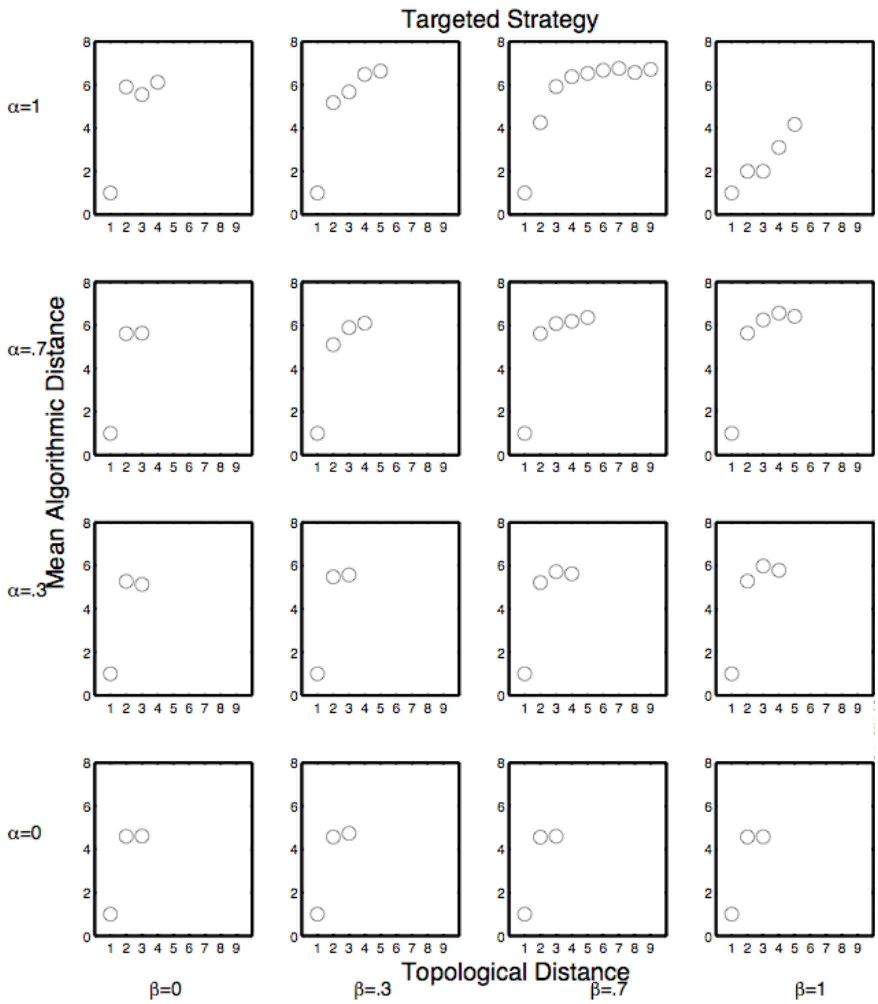

Figure 4.19 Algorithmic distance as a function of topological distance for the targeted networking strategy without attrition. Networking occurred in networks with D=3 domains and conditions of no homophily $(\alpha = 0)$ to complete homophily $(\alpha = 1)$, and no consolidation $(\beta = 0)$ to complete consolidation $(\beta = 1)$.

To summarize, there are two main obstacles in networking: attrition and the mismatch between social and network space. Attrition makes it difficult to traverse long chains and hence decreases the mean of completed chain length, but reduces success probability. When there is no attrition, the success rate increases, but the average algorithmic distance increases as well. This creates a dilemma for networkers because even though success is more likely, the longer

the algorithmic distance, the less feasible for them to reach targets. One way to solve this dilemma is by choosing an effective networking strategy and developing certain individual characteristics. Networking strategies, however, operate on social space rather than network space. Thus, the accuracy of these strategies depends, in turn, on macro structure, e.g., homophily and cross-cuttingness of social domains, and the amount of available information about targets.

There are a couple of conclusions that can be drawn from our results. First, homophily level that is too low or too high renders networking less effective. When the homophily is in the medium range, regardless of consolidation among domains, skilled actors are generally better off using either the interpersonal or targeted networking strategy. For persistent actors, however, the interpersonal strategy yields higher success rates than the targeted strategy, although success rates are not as good as for actors with perfect networking skill.

4.5 Discussions

The main finding here is that attrition has the greatest effect on searchability. There are at least three sources of attrition: lack of motivation or incentive, lack of appropriate cognitive frame, or lack of social contacts. In the context of our models, the lack of social contacts is represented in the case of extreme homophily and consolidation among social domains. This means that the macro structure is the primary driver, not individual attributes, for the darth of social capital in terms of the absence of useful social contacts. In this case

searches are terminated simply because they are disconnected components in networks. The solution to overcoming this problem is by using a networking strategy that does not depend on referrals, such as the targeted strategy. Using an appropriate networking strategy is not enough, however, because it does not guarantee that others will cooperate. Thus, this brings us to the second source of attrition, which is the lack of inclination or incentive.

We modeled individuals' lack of motivation to cooperate in a search effort as attritions that are drawn from an empirically-based attrition distribution. In our model, we use what we call skill, which is the ability to reduce others' attrition. We showed that increasing one's skill greatly increases the probability of successful searches. In practice, however, our concept of skill can be implemented in limited ways. First, the obvious way is to use power where individuals with higher social statuses or positions impose their authorities on those with lower status. Another way is to use monetary incentive. Yet, the problem of creating an efficient incentive structure for search processes is far from trivial (Kleinberg and Raghavan 2005). Individuals can also deploy various persuasion techniques (Cialdini 1998).

Another source of attrition is the lack of a cognitive frame to guide search processes. We observed this phenomena in our experiment where Target #5, who was a university professor in a large university in the Northeast, was perceived to be an easy target, so chains reaching toward this target had the lowest attrition and consequently the highest completion rate. Therefore, targets that are easy to locate in one's cognitive map will be perceived as easy to reach

and reduce attrition. In our model, actors terminate searches when they cannot find others who are closer to a target in terms of social distance. This problem is especially acute when the homophily is too low. The lower the homophily, the less relevant social categories are in the process of network formation. Therefore, there is a mismatch between the cognitive representation of social space and network space. In other words, actors are basically searching randomly.

According to our simulations, attrition stemming from the lack of a cognitive map is the worst kind, such that even actors with perfect skill using either strategy cannot achieve 100% completion rates. Although, in addition to increasing one's skill, one can also become persistent by starting many search chains. Starting as many chains as possible is a sensible strategy in this case, but of course there is a question of the ability of individuals to manage a large number of simultaneous search chains.

Another important finding is about the level of homophily that renders a search successful. Homophilies that are too low or too high are detrimental to network searchability. High homophily creates more cliques that can become disconnected to each other if the homophily is too high. However, even maximum homophily can be compensated by low consolidation among social domains. When consolidation is not too high, actors who create ties only with those who share the same category in one domain would have connections to others in different categories in another domain. Disconnected networks can only occur when homophily and consolidation are both very high.

Low homophily renders social categories irrelevant for ties formation. Because actors have only local information about network space and have to rely on their perceptions about social space to see beyond their social circles, irrelevant cognitive maps will hinder searchability. The degree of consolidation does not matter in this case; when actors make random connections in one domain, the resulting connections will also be random in all domains.

Using the above insight, organizations can take steps to ensure searchability within their organizations that can be useful, among other things, for easing knowledge transfers. Organizations can achieve medium-level homophily by maintaining formal categories based on functions, locations, or specializations and use team-based operations that comprise various individuals from different units. Thus, organizations would still have hierarchical structures but not so much for coordination tools; as for providing easily recognizable cognitive maps about who knows what in the organization.

CHAPTER FIVE: CONCLUSION

In the introduction we have argued that it is important to distinguish the topological and the algorithmic small-world hypotheses because each of them requires different empirical evidence and is relevant to different, but related, social processes. The bulk of this dissertation is about testing the algorithmic hypothesis and trying to understand the mechanisms relevant in the process of network navigation. We found partial support for the algorithmic hypothesis: there is a big difference between the estimate for mean and median of algorithmic distance. The robust median of six suggests that the algorithmic hypothesis is valid for the majority of the population. The high estimate of the mean, however, indicates that there are people who are effectively not part of the small world.

If we look at the evidence for the topological small-world hypothesis, however, we see that estimates for the mean and median are almost the same. For example, Leskovec and Horvitz (2008) found that the mean and median of the topological distance in a large online conversation network that comprises 240 million people were 6.6 and 7 respectively. Thus, comparing the evidence for the topological and algorithmic small-world hypothesis, we can conclude that the effective connectivity of social networks does not depend on topological connectivity alone. Individual strategy, motivation, and perception produce much of the variation in networking outcomes.

Generative models described in Chapter 4 allowed us to explore some candidates for the mechanism that explains how strategy, motivation, and

perception affects network outcomes. We found three mechanisms that potentially increase networking success. The first is distance reduction. Large topological distance to a target can be overcome by using the targeted networking strategy. The targeted networking strategy acts as a distance-reduction process because it allows actors to contact strangers directly, so it increases the chance of networking success because the targeted strategy renders chains short that are supposed to be long.

The second is by brute force. Here, actors' motivation (persistence) can increase networking success by starting many chains. Thus, it increases success not because it transforms long chains into short chains, but because it increases the likelihood for actors to find the closest target, the idea being because actors otherwise have no knowledge of their friends' friends and so on,

The third is attrition reduction, that is, a mechanism that increases the probability for long chains to survive through attrition reduction. In this case, actors who are capable of increasing the probability of other actors to cooperate in a networking activity have more successes because they are able to traverse longer chains. In other words, through attrition reduction, actors are able to find either nearby or far-away targets.

However, all three mechanisms described above are affected by perception. In our model, actors use one heuristic, and because the basis of heuristics is social space, then their effectiveness depends on how close the social space correlates with the network space. When homophily is absent, targets are effectively distributed randomly, and hence navigation based on the

social space becomes not very useful, i.e., actors do not have the right mental map (perception) for the network structure. Moreover, actors who can reduce attrition to zero still do not have 100% success because of perception. Actors in our model terminate a networking chain when they cannot move forward in terms of the social distance, i.e., cannot move closer to the target category or they are already in the target category but the target cannot be found. Therefore, if actors based their decision whether or not to continue a networking chain on whether, according to their perception, they can bring the chain closer to the target, then chains can terminate prematurely even when everyone cooperates.

5.1 Lessons for individuals and organizations

Based on our results, there are several ways for individuals to improve their networking abilities and for organizations to create searchable networks. In the previous chapter, we discussed three sources of attrition: lack of social contacts, lack of motivation, and lack of mental maps. The lack of social contacts occurs when we have disconnected networks as a result of extreme homophily and consolidation. Although it is not impossible, it is plausible to think that such extreme conditions rarely occur in the real world. Thus, most individuals would have difficulty in searching because of the lack of motivation or lack of mental maps.

The lack of motivation can be overcome by taking strategic actions such as using power, financial incentive, or psychological techniques to influence people so they are willing to cooperate in search processes. To reduce attrition

stemming from lack of mental maps, however, requires more subtle solutions. Individuals need to find out how people make connections so they can use that knowledge to locate a person in the cognitive social structure. One thing that we can learn from social networks studies is about the existence of multiple scales and multiple domains of social networks. To network better, individuals need to be aware of various scales that become the basis of social interactions and the multiplcity of roles and identities.

For organizations, our results suggest that having low consolidation is the best bet to ensure searchability. For any level of homophily and any networking strategy, low consolidation among social domains renders the highest probability of networking success. To achieve this, organizations first have to have more than one domain as the basis of organization structures. For example, in addition to formal organizational structure based on authority, organizations can create organization maps based on function, knowledge/specialization, or geography. Once we have several organization maps that are based on multiple domains, we can find groups of individuals who are not interacting with each other in one domain and make arrangements for them so they can interact in another domain. Organizational structures that use teams consisting of individuals with multiple functions and skills as the basic unit could lead to low consolidation and increase searchability.

On a more general level, to help individual networking better we must pay closer attention to social networks not only as a medium through which information flows, but also as a prism that individuals use to make sense of and

navigate social space (Podolny 2001). Networking is primarily driven by social space instead of network space. Once we have eliminated the motivational problem for cooperating in a networking activity, success in reaching targets depends on having some cognitive representations of social space. Thus, we should ask important questions regarding individuals' cognitive representations such as, what kind of cognitive maps should individuals use to network successfully? How accurate should the cognitive maps be? How is the construction of cognitive representations of social networks related to the social networks themselves?

5.2 Future research

There are some avenues to extend the present work. First of all, there are some relatively minor modifications to the present study that could improve our results. One way is to design the experiment such that when a participant continues a chain to a recipient, we can ask some information, e.g., demographic questions about the recipient; Milgram and Travers actually had this feature in their original experiments. Thus, even though the recipient does not respond, we can still have some basic information about them, so we have a more direct comparison between people who continued chains and those who did not. We can also modify available information about targets. For example, for the same target person we can reveal only their location but not their occupation to a group of participants, and to another group reveal their occupation but not their location, and to another group reveal both pieces of information. The idea is to see if

variations on the available information on targets would induce variation in the mental maps that guide the process and hence produce differences in search results.

It is also possible to design experiments to isolate the effects of three sources of attrition: lack of connection, lack of motivation, and lack of cognitive frame. For example, we can design experiments using data from social networking sites where we know the topological distances among actors, so we can vary other variables such as incentive or information about targets and see how they affect algorithmic distance. Another extension is to move beyond searches with known targets. Here we may need to use different designs altogether. Instead of having participants conducting a targeted search, we could ask participants to solve a problem whose solution is unknown, e.g., solving a puzzle without knowing the full picture.

There are also several possible improvements or modifications for the computational model. In our model, actors using the targeted networking strategy can make direct connections to the target category. In reality, the more specific a target group is, the harder it is to obtain access to it. Thus, there is a trade-off between easy access (but the category or group is not specific enough and hence the probability to meet a target is low) and difficult access to a very specific group in which meeting a target is very likely; for example, it's not enough to be in any party, one has to be in the "right" party and the right party is usually very exclusive.

Another extension to the model is testing the hypothesis that actors can learn by doing searches (Sabel 2004), i.e., the more an actor does a search, the better they become in searching. To do this, we need to create a dynamic model where actors can remember their previous connections and hence learn which connections are better for which search targets. Consequently, we can ponder the question that if someone can learn about doing searches, then is it possible to consider search or networking ability as a specialization; in other words, we can build networking models where there are specialized actors for doing searches. This approach may be useful for organizations. Managers in organizations are basically coordinating specialists; using this model we can study whether managers can also play the role of searching specialists.

We have argued in the introduction that the structural origin of the topological small world arises from the relational logic: two individuals who are far away in one domain are close to each other in another domain. The resulting networks are such that everyone is reasonably close to everyone else; the model in Chapter 4 formalizes this idea. Thus, the topological small world can occur in an egalitarian fashion. In the algorithmic small world, however, qualitative distinctions of individual classes matter more. In Chapter 3, we have seen that chain completion is very sensitive to individual attritions that, in turn, depend on individual attributes. Results from computer simulations also show that individual strategy, motivation, skill, and perception can render large variations in the algorithmic distance even when the underlying networks follow the topological

small-world principle. This discrepancy between the topological and algorithmic small-world distances forces us to a more nuanced “small-world” claim.

Bibliography

Adamic, L., and E. Adar. 2005. "How to search a social network." *Social Networks* 27:187-203.

Adamic, L., R. Lukose, A. Puniyani, and B. Huberman. 2001. "Search in power-law networks." *Physical Review E* 64:046135.

Andrew, Millington, Eberhardt Markus, and Wilkinson Barry. 2006. "Guanxi and supplier search mechanisms in China." *Human Relations* 59:505-531.

Barabasi, A.-L. 2002. *Linked*. Cambdrige: Perseus.

Bearman, Peter. 2005. *Doormen*. Chicago: University of Chicago Press.

Blau, P.M., and J.E. Schwartz. 1984. *Crosscutting social circles*. Orlando: Academic Press.

Boorman, Scott. 1975. "A Combinatorial Optimization Model for Transmission of Job Information Through Contact Networks." *Bell Journal of Economics* 6:216-249.

Bourdieu, P. 1986. "The forms of capital." in *Handbook for Theory and Research for the Sociology of Education*, edited by J.G. Richardson: Greenwood Press.

Cialdini, Robert. 1998. *Influence: The Psychology of Persuasion*: Collins.

Coleman, James, S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 95:S95-S120.

Dodds, Peter S., Roby Muhamad, and Duncan J. Watts. 2003. "An Experimental Study of Search in Global Social Networks." *Science* 301:827-829.

Feld, S.F. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-1035.

Ferrazzi, Keith. 2005. *Never Eat Alone: And Other Secrets to Success, One Relationship at a Time*: Broadway Business.

Geertz, Clifford. 1978. "The Bazaar Economy: Information and Search in Peasant Marketing." *American Economic Review* 68:28-32.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multileve/Hierarchical Models*. New York: Cambridge University Press.

Gibson, D.R. 2005. "Concurrency and Commitment: Network Scheduling and Its Consequences for Diffusion." *Journal of Mathematical Sociology* 29:295-323.

Gladwell, M. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. New York: Little Brown.

Goel, Sharad, Roby Muhamad, and Duncan J. Watts. 2009. "Social Search in "Small-World" Experiments." *WWW 2009*.

Granovetter, M. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-1380.

Granovetter, Mark. 1995. *Getting a Job: A Study of Contacts and Careers*. Cambridge, Mass,: Harvard University Press.

Hallen, Benjamin L., and Kathleen Eisenhardt. 2008. "How Entrepreneurs Form Relationships with Other Organizations: The Scuring of Investments from Venture Capital and Corporate Investors." in *SSRN Working Paper*. Available at SSRN: http://ssrn.com/abstract=1372328.

Hedstrom, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

Hedstrom, Peter, and Peter Bearman. 2009. *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.
Hoang, Ha, and Bostjan Antoncic. 2003. "Network-based research in entrepreneurship: A critical review." *Journal of Business Venturing* 18:165-187.
Ingram, Paul, and Michael W. Morris. 2007. "Do People Mix at Mixers? Structure, Homophily, and the "Life of the Party"." *Administrative Science Quarterly* 52:558-585.
Killworth, Peter D., Christopher McCarty, H. Russell Bernard, and Mark House. 2006. "The accuracy of small world chains in social networks." *Social Networks* 28:85-96.
Kleinberg, Jon, and Prabhakar Raghavan. 2005. "Query Incentive Networks." in *46th IEEE Symposium on Foundations of Computer Science*.
Kleinberg, Jon. 2000. "The Small-World Phenomenon: An Algorithmic Perspective." in *32nd ACM Symposium on Theory of Computing*.
Kogut, B., and G. Walker. 2001. "The Small World of Germany and The Durability of National Networks." *American Sociological Review* 66:317-335.
Korte, C., and S. Milgram. 1970. "Acquaintance Networks Between Racial Groups: Application of the Small World Method." *Journal of Personality and Social Psychology* 15:101-108.
Kossinets, G., and D. J. Watts. 2006. "Empirical Analysis of an Evolving Social Network." *Science* 311:88-90.
Kossinets, Gueorgi., and Duncan Watts, J. 2005. "Recovery of social network structure from discrete communication data." Columbia University.
Lee, John B., and Duncan J. Watts. 2006. "When Does Networking Work?" in *Working paper. ISERP Columbia University*.
Lee, Nancy, H. 1969. *The Search for an Abortionist*. Chicago: The University of Chicago Press.
Leskovec, J., and E. Horvitz. 2008. "Planetary-scale views on a large instant-messaging network." *Proceedings of the 17th International World Wide Web Conference*.
Lin, N. 2001. *Social capital: a theory of social structure and action*: Cambridge University Press.
Martin, John, L. 2009. *Social Structures*. Princeton: Princeton University Press.
Montgomery, James. 1994. "Weak Ties, Employment, and Inequality: An Equilibrium Analysis." *American Journal of Sociology* 99:1212-1236.
Motter, A.E., T. Nishikawa, and Y. Lai. 2003. "Large-scale structural organization of social networks." *Physical Review E* 68.
MWAAW, Messaging Anti-Abuse Working Group. 2006. "Email Metrics Program: The Network Operators' Perspective." in *Email Metrics Report - Second Quarter 2006*.
Nohria, Nitin. 1992. "Information and Search in the Creation of New Business Ventures: The Case of the 128 Venture Group." in *Networks and Organizations: structure, form, and action*, edited by Nitin Nohria and Robert Eccles. Cambridge: Harvard University Press.
Podolny, Joel M. 2001. "Networks as the Pipes and Prisms of the Market." *American Journal of Sociology* 107:33-60.

Raider, Holly J., and Ronald Burt, S. 1996. "Boundaryless Careers and Social Capital." Pp. 187-200 in *The Boundaryless Career: A New Employment Principle for a New Organizational Era*, edited by M.B. Arthur and D.M. Rousseau. New York: Oxford University Press.

Renzulli, Linda, Howard Aldrich, and Jame Moody. 2000. "Family Matters: Gender, Networks, and Entrepreneurial Outcomes." *Social Forces* 79:523-546.

Sabel, Charles F. 2004. "A real time revolution in routines." in *The firm as a collaborative community*, edited by C. Heckscher and P. Adler: Oxford University Press.

Simmel, Georg. 1955. *Conflict and The Web of Group-Affiliations*. New York: The Free Press.

Singh, Jasjit, Morten T. Hansen, and Joel M. Podolny. 2008. "The World is Not Small for Everyone: Pathways of Discrimination in Searching for Information in Organizations." Singapore: INSEAD.

Small, Mario, L. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*: Oxford University Press.

Sorenson, Olav, and Toby E. Stuart. 2008. "Bringing The Context Back In: Settings and The Search for Syndicate Partners in Venture Capital Investment Network." *Administrative Science Quarterly* 53:266-294.

Stark, David. 2009. *The Sense of Dissonance: Accounts of Worh in Economic Life*: Princeton University Press.

Stuart, Toby E., and Waverly Ding. 2006. "When do Scientists Become Entrepreneurs? The Social Structure Antecedents of Commercial Activity in the Academic Life Sciences." *American Journal of Sociology* 112:97-144.

Stuart, Toby E., and Olav Sorensen. 2008. "Strategic networks and entrepreneurial ventures." *Strategic Entrepreneurship Journal* 1:211-227.

Travers, J., and Stanley. Milgram. 1969. "An Experimental Study of the Small World Problem." *Sociometry* 32:425-443.

Uzzi, Brian, and Shannon Dunlap. 2005. "How To Build a Better Network." *Harvard Business Review* 83:53-60.

Vissa, Balagopal. 2008. "Bringing the Individual Back In: Entrepreneurs' Networking Actions and Referral Based Search for New Exchange Partners." in *Working paper*. Singapore: INSEAD.

Watts, D. J. 2003. *Six Degrees: The Science of A Connected Age*. New York: W.W.Norton.

Watts, D.J., and S.H. Strogatz. 1998. "Collective Dynamics of "Small-World" Networks." *Nature* 393:440-442.

Watts, Duncan J. 1999. "Networks, Dynamics, and the Small-World Phenomenon." *American Journal of Sociology* 105:493-527.

Watts, Duncan J., Peter S. Dodds, and Mark E.J. Newman. 2002. "Identity and search in social networks." *Science* 296:1302-1305.

White, Harrison, C. 1970. "Search Parameters for the Small World Problem." *Social Forces* 49:259-264.

—. 1992. *Identity and Control*. Princeton, NJ: Princeton University Press.

Zott, C., and Q.N. Huy. 2007. "How entrepreneurs use symbolic management to acquire resources." *Administrative Science Quarterly* 52:70-105.

Appendix: Questionnaires

Experiment 1
Individual attributes:

1. **Name**: First name, last name.
2. **E-mail address**.
3. **Gender**: Female, Male
4. **Age**: 18-29, 30-39, 40-49, 50-59, >60.
5. **Education level**: Elementary School, High School, College/University, Graduate School.
6. **Income level**: <$2,000, $2,000 - $25,000, $25,000 - $49,000, $50,000-$100,000, >$100,000.
7. **Occupation**: Military or Police,Banking or Finance,Government or Local Administration,Arts or Culture,Information Technology,Health,Science,Education,Consumer Services,Law,Industry,Advertising,Religion,Agriculture,Media,Sports,Construction,Telecommunications,Commerce,Transportation,Tourism,Community Service,Parent,Other.
8. **Work Position**: Entrepreneur,Chief Executive,Executive,Manager,Specialist or Engineer,Technical Personnel,Administrative Personnel,Freelance,Skilled Worker,Graduate Student,College Student,High School Student,Housekeeper,Unemployed,Retired,Other.
9. **Religion**: Christianity, None, Judaism, Hindu, Buddhism, Islam, Other.
10. **Ethnic category**: White or Caucasian,Hispanic,Black or African,Asian,South Asian,South East Asian,Central Asian,Middle Eastern,Indigenous,Mixed Race,Pacific Islander,Other,No Answer.
11. **Country**: AF,AL,DZ,AS,AD,AO,AI,AQ,AG,AR,AM,AW,AU,AT,AZ,BS,BH,BD,BB,BY,BE,BZ,BJ,BM,BT,BO,BA,BW,BV,BR,IO,BN,BG,BF,BI,KH,CM,CA,CV,KY,CF,TD,CL,CN,CX,CC,CO,KM,CG,CD,CK,CR,CI,HR,CU,CY,CZ,DK,DJ,DM,DO,TP,EC,EG,SV,GQ,ER,EE,ET,FK,FO,FJ,FI,FR,GF,PF,TF,GA,GM,GE,DE,GH,GI,GR,GL,GD,GP,GU,GT,GN,GW,GY,HT,HM,HN,HK,HU,IS,IN,ID,IR,IQ,IE,IL,IT,JM,JP,JO,KZ,KE,KI,KR,KP,KW,KG,LA,LV,LB,LS,LR,LY,LI,LT,LU,MO,MK,MG,MW,MY,MV,ML,MT,MH,MQ,MR,MU,YT,MX,FM,MD,MC,MN,MS,MA,MZ,MM,NA,NR,NP,AN,NL,NC,NZ,NI,NE,NG,NU,NF,MP,NO,OM,PK,PW,PA,PG,PY,PE,PH,PN,PL,PT,PR,QA,RE,RO,RU,RW,SH,KN,LC,PM,VC,WS,SM,ST,SA,SN,SC,SL,SG,SK,SI,SB,SO,ZA,GS,ES,LK,SD,SR,SJ,SZ,SE,CH,SY,TW,TJ,TZ,TH,TG,TK,TO,TT,TN,TR,TM,TC,TV,UG,UA,AE,UK,US,UM,UY,UZ,VU,VA,VE,VN,VG,VI,WF,YE,YU,ZM,ZW.
12. **City**: open values.
13. **Send results**: Yes, No.

Relational variables:

1. **Strength of relationship**: Extremely Close,Very Close,Fairly Close,Casually,Not close.
2. **Origin of relationship**: Immediate Family,Extended Family,Friend of Family,Grew up together,Mutual Friend,School,Work,Party/Bar/Cafe,Live in same neighborhood,Travel/Exchange/Penpal,Faith/Volunteering,Hobby/Sport/Interest,Internet,Other.
3. **Nature of relationship**: Parent,Child,Sibling,Relative,In-law,Spouse or Significant Other,Ex,Friend,Coworker (Unspecified),Senior Coworker,Junior Coworker,Teacher,Student,Client,Service Provider,Spiritual Guide,Other.
4. **Reason for choosing recipient**: home geography,family origin geography,travel geography,similar profession,similar education,similar religion,work brings contact,lots of friends,will continue chain,is or knows target,no reason or hunch,same last name,other.

Experiment 2

Individual attributes:

1. **Name**: First name, last name.
2. **E-mail address**.
3. **Country**: ”Afghanistan”, ”Albania”, ”American Samoa”, ”Angola”, ”Antarctica”, ”Antigua And Barbuda”, ”Argentina”, ”Armenia”, ”Australia”, ”Austria”, ”Bahamas, ”Bahrain”, ”Bangladesh”, ”Barbados”, ”Belarus”, ”Belgium”, ”Bermuda”, ”Bolivia”, ”Bosnia and Herzegovina”, ”Brazil”, ”Brunei”, ”Bulgaria”, ”Cameroon”, ”Canada”, ”Cayman Islands”, ”Chile”, ”China”, ”Cocos (Keeling) Islands”, ”Colombia”, ”Congo”, ”Costa Rica”, ”Croatia (Hrvatska)”, ”Cuba”, ”Cyprus”, ”Czech Republic”, ”Denmark”, ”Dominican Republic”, ”East Timor”, ”Ecuador”, ”Egypt”, ”Eritrea”, ”Estonia”, ”Ethiopia”, ”Faroe Islands”, ”Finland”, ”France”, ”French Guiana”, ”Gabon”, ”Germany”, ”Greece”, ”Guam”, ”Guatemala”, ”Guyana”, ”Honduras”, ”Hongkong S.A.R”, ”Hungary”, ”Iceland”, ”India”, ”Indonesia”, ”Iran”, ”Ireland”, ”Israel”, ”Italy”, ”Jamaica”, ”Japan”, ”Kazakhstan”, ”Kenya”, ”Korea”, ”Kuwait”, ”Kyrgyzstan”, ”Laos”, ”Latvia”, ”Lebanon”, ”Libya”, ”Luxembourg”, ”Macedonia. Former Yugoslav Rep”, ”Malaysia”, ”Malta”, ”Martinique”, ”Mauritius”, ”Mexico”, ”Monaco”, ”Morocco”, ”Mozambique”, ”Nepal”, ”Netherlands, ”Netherlands Antilles”, ”New Zealand”, ”Nicaragua”, ”Nigeria”, ”Northern Mariana Islands”, ”Norway”, ”Oman”, ”Pakistan”, ”Panama”, ”Paraguay”, ”Peru”, ”Philippines”, ”Poland”, ”Portugal”, ”Puerto Rico”, ”Qatar”, ”Reunion”, ”Romania”, ”Russia”, ”Saint Kitts And Nevis”, ”Saint Lucia”, ”Saint Vincent And The Grenadin”, ”Saudi Arabia”, ”Singapore”, ”Slovakia”, ”Slovenia”, ”South Africa”, ”Spain”, ”Sri Lanka”, ”Sweden”, ”Switzerland”, ”Taiwan”, ”Tajikistan”, ”Tanzania”, ”Thailand”, ”Trinidad And Tobago”, ”Turkey”, ”Uganda”, ”Ukraine”, ”United Arab Emirates”, ”United Kingdom”, ”United States”, ”United States Minor Outlying I”, ”Uruguay”, ”Vanuatu”, ”Venezuela”, ”Vietnam”, ”Virgin Islands (US)”, ”Wallis And Futuna Islands”, ”Yugoslavia”, ”Zimbabwe”.
4. **City**: open values.

5. **State** : “Non US/Canada resident”Alabama”, ”Alaska”, ”Alberta”, ”Arizona”, ”Arkansas”, ”British Columbia”, ”California”, ”Colorado”, ”Connecticut”, ”Delaware”, ”District of Columbia”, ”Florida”, ”Georgia”, ”Hawaii”, ”Idaho”, ”Illinois”, ”Indiana”, ”Iowa”, ”Kansas”, ”Kentucky”, ”Louisiana”, ”Maine”, ”Manitoba”, ”Maryland”, ”Massachusetts”, ”Michigan”, ”Minnesota”, ”Mississippi”, ”Missouri”, ”Montana”, ”Nebraska”, ”Nevada”, ”New Brunswick”, ”Newfoundland”, ”New Hampshire”, ”New Jersey”, ”New Mexico”, ”New York”, ”Non US/Canada Resident”, ”North Carolina”, ”North Dakota”, ”Northwest Territories”, ”Nova Scotia”, ”Ohio”, ”Oklahoma”, ”Ontario”, ”Oregon”, ”Pennsylvania”, ”Prince Edward”, ”Quebec”, ”Saskatchewan”, ”South Carolina”, ”South Dakota”, ”Tennessee”, ”Texas”, ”US Military: AA”, ”US Military: AE”, ”US Military: AP”, ”Utah”, ”Vermont”, ”Virginia”, ”Virgin Islands”, ”Washington”, ”West Virginia”, ”Wisconsin”, ”Wyoming”, ”Yukon Territories”.
6. **Zip/Postal Code**: open values.
7. **Q: In which country were you born?**: Valid values are the same as countries.
8. **Q: How long have you lived in your current neighborhood?**: ”0-1 year”, ”1-2 years”, ”3-5 years”, ”6-9 years”, ”10-19 years”, ”20+ years”.
9. **Marital Status**: ”Married”, ”Living with”, ”Single(never married)”, ”Divorced/Sep”, ”Widowed”.
10. **Ethnic categories**: ”Asian”, ”Black or Africa”, ”Central Asian”, ”European”, ”Hispanic”, ”Indigenous”, ”Middle Eastern”, ”Other”, ”Pacific Islande”, ”South Asian”, ”South East Asia”, ”White or Caucasian”.
11. **Gender**: Female, Male.
12. **Age**: ”0-12”, ”13-18”, ”18-24”, ”25-29”, ”30-34”, ”35-39”, ”40-44”, ”45-49”, ”50-54”, ”55-59”, ”60-64”, ”65-69”, ”70-79”, ”80+”.
13. **Q: Industry in which you work**? : "Accounting", "Agriculture/Farming", "Architecture/Design", "Arts/Entertainment", "Computers/Software/Technology", "Construction", "Consulting", "Education/Schools/Academia", "Energy/Utilities/Fuel/Chemicals", "Engineering", "Finance/Banking/Brokerage", "Government/Diplomatic services", "Health Care/Hospitals", "Import/Export/Trade", "Information Management/Library", "Insurance", "Legal", "Manufacturing", "Marketing/Advertising/Communications/PR", "Media/Publishing/Broadcasting", "Military", "Non-profit/Associations", "Pharmaceuticals", "Real Estate/Property Management", "Recruiting/Staffing/Human Resources", "Religious Institutions", "Research & Development/Research", "Retail", "Social Services", "Telecommunications", "Transportation", "Travel/Hospitality/Service", "Wholesale", "Homemaker", "Student", "Retired", "Other".
14. **Job title**: "Accountant/Auditor", "Administrative Assistant", "Analyst", "Artist/Musician/Actor/Entertainer", "Architect", "Associate", "Broker/Trader/Advisor", "CEO/President/Chairman", "CFO, COO, CTO, CIO, CMO", "Clergy", "Clerical worker", "Computer professional", "Consultant", "Director", "Educator/Teacher/Professor", "Engineer", "Entrepreneur",

"Government official", "Health care worker (other than doctor)", "Homemaker", "Lawyer/Judge", "Manager", "Military Officer", "Partner/Principal/Owner", "Researcher", "Sales Manager/Account Executive", "Skilled laborer", "Scientist", "Service provider", "Student", "Supervisor", "Technician", "Volunteer", "Vice President/SVP/EVP", "Writer/Editor", "Retired", "Other".

15. **Religion** : ”Atheism”, ”Buddhism”, ”Christianity”, ”Hinduism”, ”Islam”, ”Judaism”, ”None”, ”Other”.
16. **First Language**: "English", "Spanish", "French", "Afrikaans", "Albanian", "Amharic", "Arabic", "Armenian", "Assamese", "Azeri", "Balinese", "Basque", "Bengali", Bhojpuri", "Bikol", "Bosnian", "Bulgarian", "Burmese", "Cantonese", "Catalan", "Cebuano", "Chinese", "Croatian", "Czech", "Danish", "Dari", "Dutch", "Estonian", "Farsi", "Finnish", "Flemish", "French", "Fuzhou", "Ga", "Georgian", "German", "Greek", "Gujarati", "Haitian Creole", "Hakka", "Hausa", "Hebrew", "Hiligaynon", "Hindi", "Hmong", "Hokkien", "Hungarian", "Icelandic", "Ilocano", "Indonesian", "Italian", "Japanese", "Javanese", "Kannada", "Kazakh", "Khmer", "Korean", "Kurdish", "Lao", "Latvian", "Lingala", "Lithuanian", "Luganda", "Macedonian", "Malay", "Malayalam", "Mandarin", "Maori", "Marathi", "Mongolian", "Ndebele", "Nepali", "Norwegian", "Oriya", "Pashto", "Polish", "Portuguese", "Punjabi", "Quechua", "Romanian", "Russian", "Serbian", "Sindhi", "Sinhalese", "Slovak", "Slovenian", "Somali", "Spanish", "Sundanese", "Swahili", "Swedish", "Tagalog", "Tamil", "Tatar", "Telugu", "Thai", "Tibetan", "Tigrinya", "Turkish", "Turkmen", "Ukrainian", "Urdu", "Uzbek", "Vietnamese", "Waray", "Xhosa", "Yoruba", "Zulu", "Other".
17. **Highest level of school completed**: ”Completed College/University (e.g., a bachelors degree”, ”Completed Elementary/Primary School/6 years of Sch”, ”Completed High School/12 years of School”, ”Completed Junior High School/9 years of School”, ”Doctorate degree”, ”Masters degree or equivalent”, ”No Schooling completed”, ”Professional degree (e.g. Medicine, Law)”, ”Some Elementary/Primary School completed”, ”Some time at College/University, no degree” .
18. **Q: Where is your computer located?**: ”Public Internet Access (e.g. Internet cafe, Library)”, ”Home”, ”Office/School/University”.
19. **Annual income**: ”Much lower than the average”, ”Much higher than the average”, ”Higher than the average”, ”Lower than the average”, ”Around the average”.
20. **Q. Do you want to be notified when experiment results become available?** Yes, No.

Relational variables:

1. **What is the nature of your relationship?** “Spouse/Partner/Significant Other”, “Family member (e.g., father, aunt, cousin)”, ”Friend/Acquaintance”, “Coworker/Professional Colleague”, “Junior Colleague”, “Senior Colleague”, “Employer”, “Employee”, “Teacher/Professor/Instructor”, “Student/Pupil”, “Business Partner”, “Religious/Community leader”, “Customer/Client”, “Service provider (e.g., doctor, lawyer)”, “Other.”

2. **How did you get to know them?** “Belong to same family”, “Profession/Work”, “School/College/University”, “Social gathering/organization (e.g., party, nightclub)”, “Political gathering/organization”, “Professional gathering/organization”, “Internet (e.g., chat room, mailing list)”, “Place of worship or religious activities”, “Live(d) in the same neighborhood”, “Through shared sporting interests”, “Through shared hobby/club/entertainment”, “Do not remember”, “Other.”
3. **Were you introduced by a mutual friend?** No, Yes.
4. **How well do you know this person?** “Extremely Close”, “Very Close”, “Close”, “Reasonably Close”, “Somewhat Close”, “Not Close.”
5. **Approximately how often do you communicate using email with this person?** “Many times per day”, “Daily”, “Weekly”, “Less Often”, “Never.”
6. **Would you socialize with this person?** (e.g., have lunch or see a movie together)? No, Yes.
7. **Would you ask this person for advice?** (e.g., on choosing a job/school)? No, Yes.
8. **Would you feel comfortable asking this person for a favor?** (e.g., helping you move house, looking after a pet while youre away). No, Yes.
9. **Would you discuss personal matters with this person?** (e.g., an illness, a family dispute). No, Yes.
10. **Would you be comfortable lending money to this person?** (a week of your wages or more). No, Yes.
11. **We would like to know why you chose this person. What is it about them that will help move your message ”closer” to the target?** “Where they live”, “Where they used to live”, “Where they have traveled to”, “Their nationality”, “They know someone who matches one or more of the above geography categories”, “The company or organization they work for”, “The company or organization they used to work for”, “Their profession”, “Their former profession”, “They know someone who matches one or more of the above work categories”, “Their schooling/education”, “Their community involvement”, “Their religious affiliations”, “Their hobbies or interests”, “Their race or ethnicity”, “They know someone who matches one or more of the above personal categories”, “They know many people”, “They know many different types of people”, “They are likely to pass this message on”, “They know someone who matches one or more of the above network categories”, “Other.”